

VŠB – Technická univerzita Ostrava  
Fakulta elektrotechniky a informatiky  
Katedra informatiky

Nástroj pro analýzu a budování  
zpětných odkazů

Tools for building and evaluation of  
back links

2014

Tomáš Pasičnyk

## Zadání diplomové práce

Student: **Bc. Tomáš Pasičnyk**

Studijní program: N2647 Informační a komunikační technologie

Studijní obor: 2612T025 Informatika a výpočetní technika

Téma: **Nástroj pro analýzu a budování zpětných odkazů**  
**Tools for building and evaluation of back links**

### Zásady pro vypracování:

Cílem diplomové práce bude nástroj pro efektivní správu zpětných odkazů (tzv. link building). Součástí práce bude podrobná analýza efektivy budování zpětných odkazů u nového a dlouhodobě existujícího webu se závěrečným zhodnocením kvality a efektivy volby zpětných odkazů jejich udržování a rozšiřování.

1. Seznamte se a analyzujte současné trendy v oblasti SEO a linkbuildingu.
2. Seznamte se s algoritmy vyhledávacích enginů, které hodnotí webové stránky a zpětné odkazy.
3. Seznamte se s existujícími řešeními (WebCEO, Link Assistant, SEOnastroje.cz).
4. Testujte různé techniky linkbuildingu s vyhodnocením účinnosti.
5. Definujte způsoby pro budování zpětných odkazů (diskuzní fóra, blogy, PR články, sociální sítě, katalogy, partnerské weby, placené odkazy).
- 6 Implementujte nástroje pro budování, správu a testování přínosu zpětných odkazů:
  - a) aplikace bude sloužit tvůrcům webových stránek, ke zvýšení kvality a hodnocení webu (PageRank, pozice ve výsledcích vyhledávání, atd.),
  - b) aplikace vyhledá zpětné odkazy daného webu a analyzuje výsledky v porovnání s konkurenčními weby,
  - c) vyhledá webové stránky s podobnou tematikou,
  - d) vyhledá a analyzuje konkurenční weby za účelem zjištění jejich zpětných odkazů,
  - e) aplikace bude doporučovat oblasti/weby, ve kterých je vhodné budovat zpětné odkazy a techniky pro jejich vkládání,
  - f) aplikace se bude učit ze získaných dat,
  - g) bude sloužit jako manažer všech zpětných odkazů umístěných v různých oblastech (blogy, sociální sítě, katalogy, PR weby atd.).
7. Testujte aplikaci na reálných webových stránkách a vyhodnoťte dosažené výsledky.
8. Výsledky implementace i analýzy ověřte v praxi na reálných projektech.

### Seznam doporučené odborné literatury:

- [1] KUBÍČEK, Michal. Velký průvodce SEO: jak dosáhnout nejlepších pozic ve vyhledávacích. Vyd. 1. Brno: Computer Press, 2008, 318 s. ISBN 978-80-251-2195-5.
- [2] KUBÍČEK, Michal a Jan LINHART. 333 tipů a triků pro SEO: [sbírka nejlepších technik optimalizace webů pro vyhledávače]. Vyd. 1. Brno: Computer Press, 2010, 262 s. ISBN 978-80-251-2468-0.
- [3] GRAPPONE, Jennifer a Grativa COUZIN. SEO: search engine optimization : ovládněte SEO a získejte výhodu před konkurencí : optimalizujte své webové stránky pro vyhledávací servery : přiveďte na své stránky zákazníky dříve, než to udělá konkurence. Vyd. 1. Brno: Computer Press, 2010, 262 s. ISBN 978-80-86815-85-5.

[4] DOMES, Martin a Gradiva COUZIN. SEO: jednoduše. Vyd. 1. Překlad Roman Skřivánek, Dana Balaščíková. Brno: Computer Press, 2011, 141 s. Naučte se za víkend (Computer Press). ISBN 978-80-251-3456-6.

[5] JANOUGH, Viktor a Gradiva COUZIN. Internetový marketing: prosad'te se na webu a sociálních sítích. Vyd. 1. Překlad Roman Skřivánek, Dana Balaščíková. Brno: Computer Press, 2010, 304 s. Naučte se za víkend (Computer Press). ISBN 978-80-251-2795-7.

Formální náležitosti a rozsah diplomové práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí diplomové práce: **Ing. Radoslav Fasuga, Ph.D.**

Datum zadání: 01.09.2013

Datum odevzdání: 07.05.2014



doc. Dr. Ing. Eduard Sojka  
vedoucí katedry



prof. RNDr. Václav Snášel, CSc.  
děkan fakulty

## Prohlášení studenta

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně.

Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

Dne: 29. dubna 2014

*Pasický Tomáš*

Podpis

## Poděkování

Rád bych poděkoval vedoucímu diplomové práce Ing. Radoslavu Fasugovi, Ph.D. za odbornou pomoc a konzultaci při vytváření této práce.

# Abstrakt

Úvod diplomové práce popisuje optimalizaci internetových stránek pro vyhledávače. Tato část se kromě on-page faktorů věnuje především procesu budování zpětných odkazů s popisem oblastí a technik pro jejich vkládání. Dále jsou zkoumány algoritmy vyhledávacích enginů Google a Seznam, které mají na starost hodnocení webových stránek. Následující kapitola obsahuje popis a srovnání existujících nástrojů na SEO analýzu se zaměřením na funkce související se zpětnými odkazy. Jedná se o Collabim, SEO Nástroje, LinkAssistant, SEO SpyGlass a Web CEO. Hlavní část textu je věnována popisu vlastního nástroje sloužícího k analýze odkazového profilu webu a podpoře při budování zpětných odkazů a také modulu pro stahování aukcí z portálu Aukro. Obsahuje popis funkcí obou systémů a postupy při jejich implementaci. V závěru jsou shrnuty výsledky dosažené během testování aplikace a zhodnocení jejich přínosu.

## Klíčová slova

zpětný odkaz, SEO, vyhledávače, klíčové slovo, odkazový profil, budování zpětných odkazů, Aukro, Google

## Abstract

Introduction of the thesis describes the optimization of websites for search engines. This section primarily describes the process of link-building with description of areas and techniques for their publishing. Furthermore, algorithms of search engines Google and Seznam which affect the ranking of web pages are studied. The following chapter contains a description and comparison of existing tools for SEO analysis, focusing on the functions associated with link-building. This is a Collabim, SEO Tools, LinkAssistant, SEO SpyGlass and Web CEO. The main part of the text is focused on describing own tool to analyze link profile and support in link-building and a module for download auctions from Aukro. It contains a description of these two systems and process of their implementation. The conclusion summarizes the results obtained during application testing and evaluation of their benefits.

## Key words

back link, SEO, search engines, keyword, link profile, link building, Aukro, Google

# Seznam použitých symbolů a zkratek

**CSS** - *Cascading Style Sheets* – jazyk pro úpravu vzhledu HTML dokumentů

**JSON** – *JavaScript Object Notation* – způsob zápisu dat nezávislý na platformě

**PPC** – *Pay per Click*

**SERP** – *Search Engine Results Page* – stránka s výsledky vyhledávání

**API** - *Application Programming Interface*

**RPC** - *Remote procedure call*

**Cron** – systémový proces spouštěný automaticky

**CSV** - *Comma-Separated Values* – souborový formát

**ČR** – *Česká republika*

**DOM** – *Document Object Model* – objektová reprezentace XML nebo HTML dokumentu

**GNU/GPL** – typ softwarové licence

**HTML** – *HyperText Markup Language* - značkovací jazyk

**IČO (IČ)** – *identifikační číslo organizace*

**ID** – *identifikační číslo*

**IP** – *Internet Protocol*

**MB** - *megabyte*

**MVC** – *Model-View-Controller*

**MySQL** – typ SŘBD

**PHP** – *Hypertext Preprocessor* - programovací jazyk pro tvorbu webových aplikací

**SEO** – *Search Engine Optimization*

**SOAP** - *Simple Object Access Protocol*

**SŘBD** – *systém řízení báze dat*

**URL** - *Uniform Resource Locator*

**XML** - *Extensible Markup Language* – značkovací jazyk

## Seznam tabulek

Tabulka 1: Rozdíl mezi skutečným PR a GTPR .....	- 13 -
Tabulka 2: Ceník tarifů Collabim .....	- 19 -
Tabulka 3: Tarify a omezení u Web CEO .....	- 28 -
Tabulka 4: Datový slovník tabulky bac_backlinkSite .....	- 32 -
Tabulka 5: Srovnání výsledků pro web ceske-mince.cz .....	- 63 -
Tabulka 6: Srovnání výsledků pro web podlahy-chylek.cz .....	- 63 -
Tabulka 7: Souhrn analýz .....	- 64 -
Tabulka 8: Časy stahování .....	- 65 -
Tabulka 9: Parametry serveru .....	- 65 -
Tabulka 10: Souhrn stahování aukcí .....	- 68 -

## Seznam obrázků

Obrázek 1: Graf obecné distribuce Page Ranku .....	- 12 -
Obrázek 2: HITS algoritmus .....	- 15 -
Obrázek 3: Graf použití vyhledávačů v ČR - týden 15/2014 .....	- 16 -
Obrázek 4: Ukázka nástroje Collabim .....	- 17 -
Obrázek 5: Ukázka evidence odkazů nástroje Collabim .....	- 18 -
Obrázek 6: Ukázka nástroje SEOnástroje.cz .....	- 20 -
Obrázek 7: Ukázka části reportu nástroje LinkAssistant .....	- 23 -
Obrázek 8: Ukázka nástroje Seo SpyGlass .....	- 24 -
Obrázek 9: Ukázka reportu SEO SpyGlass .....	- 25 -
Obrázek 10: Ukázka nástroje Web CEO – shrnutí analýzy odkazů .....	- 27 -
Obrázek 11: ER-diagram databáze modulu .....	- 33 -
Obrázek 12: Vývojový diagram nalezení zpětného odkazu .....	- 38 -
Obrázek 13: Přehled uživatelských projektů .....	- 42 -



Obrázek 14: Ukázka shrnutí odkazových profilů u svých webů .....	- 42 -
Obrázek 15: Ukázka srovnání s konkurenčními weby .....	- 43 -
Obrázek 16: Ukázka posledních změn odkazového profilu u projektových webů .....	- 44 -
Obrázek 17: Ukázka výpisu odkazujících domén .....	- 44 -
Obrázek 18: Ukázka seznamu textů odkazů.....	- 45 -
Obrázek 19: Ukázka výpisu odkazujících stránek .....	- 46 -
Obrázek 20: Ukázka grafu odkazujících TLD .....	- 47 -
Obrázek 21: Ukázka grafu typů odkazů .....	- 47 -
Obrázek 22: Ukázka kontroly důležitých domén .....	- 48 -
Obrázek 23: Seznam katalogů v plánu .....	- 50 -
Obrázek 24: Diagram výpočtu hodnoty katalogu.....	- 52 -
Obrázek 25: Ukázka výběru tematických webů.....	- 53 -
Obrázek 26: API export dat o projektu .....	- 54 -
Obrázek 27: API - import/export katalogů.....	- 55 -
Obrázek 28: API - export katalogů k registraci.....	- 56 -
Obrázek 29: API - update seznamu katalogů k registraci .....	- 56 -
Obrázek 30: Seznam ukončených aukcí produktu .....	- 57 -
Obrázek 31: Vyhledávací fráze produktu pro Aukro.cz.....	- 58 -
Obrázek 32: ER-diagram databáze Aukro modulu .....	- 60 -
Obrázek 33: Graf vývoje pozic klíčových slov v Google .....	- 62 -
Obrázek 34: Graf pozic klíčové fráze "renovace podlah" .....	- 66 -
Obrázek 35: Graf pozic fráze "pokládka podlah frýdek" .....	- 67 -

## Seznam rovnic

Rovnice 1: Obecná rovnice pro výpočet Page Ranku .....	- 12 -
Rovnice 2: Vážený průměr webu .....	- 43 -
Rovnice 3: Výpočet hodnoty odkazu .....	- 45 -

# Obsah

<b>1</b>	<b>ÚVOD .....</b>	<b>- 1 -</b>
<b>2</b>	<b>SEARCH ENGINE OPTIMIZATION (SEO) .....</b>	<b>- 2 -</b>
<b>2.1</b>	<b>Co je to SEO? .....</b>	<b>- 2 -</b>
<b>2.2</b>	<b>SEO faktory .....</b>	<b>- 2 -</b>
2.2.1	On-page .....	- 2 -
2.2.2	Off-page .....	- 5 -
2.2.3	Black Hat SEO .....	- 5 -
<b>2.3</b>	<b>Budování zpětných odkazů .....</b>	<b>- 7 -</b>
2.3.1	Důležitost zpětných odkazů .....	- 7 -
2.3.2	Techniky budování zpětných odkazů .....	- 7 -
2.3.3	Definování oblastí pro budování zpětných odkazů .....	- 8 -
<b>2.4</b>	<b>Vyhledávací enginy a jejich algoritmy .....</b>	<b>- 11 -</b>
2.4.1	Google .....	- 11 -
2.4.2	Seznam .....	- 14 -
2.4.3	Srovnání .....	- 16 -
<b>3</b>	<b>EXISTUJÍCÍ ŘEŠENÍ PRO SEO ANALÝZU .....</b>	<b>- 17 -</b>
<b>3.1</b>	<b>Collabim .....</b>	<b>- 17 -</b>
3.1.1	Funkce .....	- 17 -
3.1.2	Ceník tarifů .....	- 19 -
<b>3.2</b>	<b>SeoNástroje .....</b>	<b>- 20 -</b>
<b>3.3</b>	<b>SEO Power Suite .....</b>	<b>- 22 -</b>
3.3.1	LinkAssistant .....	- 22 -
3.3.2	Seo SpyGlass .....	- 23 -
<b>3.4</b>	<b>WebCEO .....</b>	<b>- 26 -</b>
3.4.1	Funkce .....	- 26 -
3.4.2	Tarify a omezení .....	- 28 -
<b>3.5</b>	<b>Shrnutí .....</b>	<b>- 29 -</b>

<b>4</b>	<b>NÁSTROJ PRO ANALÝZU A BUDOVÁNÍ ODKAZOVÉHO PROFILU.....</b>	<b>- 30 -</b>
<b>4.1</b>	<b>Funkční analýza .....</b>	<b>- 30 -</b>
4.1.1	Funkce systému .....	- 30 -
4.1.2	Vstupní data .....	- 30 -
4.1.3	Výstupy ze systému .....	- 31 -
4.1.4	Uživatelé systému .....	- 31 -
<b>4.2</b>	<b>Databázový model.....</b>	<b>- 31 -</b>
4.2.1	Datový slovník .....	- 32 -
4.2.2	ER-Diagram.....	- 33 -
<b>4.3</b>	<b>Popis implementace .....</b>	<b>- 34 -</b>
4.3.1	Proces hledání zpětných odkazů .....	- 34 -
4.3.2	Parsování odkazujících stránek .....	- 37 -
4.3.3	Aktualizace zpětných odkazů .....	- 40 -
4.3.4	Optimalizace procesů .....	- 40 -
4.3.5	Použité technologie.....	- 41 -
4.3.6	Nasazení .....	- 41 -
<b>4.4</b>	<b>Popis funkcí systému .....</b>	<b>- 42 -</b>
4.4.1	Správa projektu .....	- 42 -
4.4.2	Analýza odkazového profilu .....	- 42 -
4.4.3	Budování zpětných odkazů.....	- 49 -
4.4.4	Využití databáze znalostí.....	- 51 -
4.4.5	Vyhledávač tematických webů.....	- 52 -
4.4.6	Automatizace procesů.....	- 53 -
4.4.7	Administrační rozhraní .....	- 53 -
4.4.8	API .....	- 54 -
<b>5</b>	<b>MODUL PRO STAHOVÁNÍ AUKCÍ Z PORTÁLU AUKRO.CZ.....</b>	<b>- 57 -</b>
<b>5.1</b>	<b>Popis funkcí modulu .....</b>	<b>- 57 -</b>
5.1.1	Zobrazení aukcí u produktu.....	- 57 -
<b>5.2</b>	<b>Popis implementace .....</b>	<b>- 58 -</b>
5.2.1	Proces vyhledání aukce .....	- 58 -
5.2.2	Aukro WebAPI .....	- 59 -
5.2.3	Automatizace procesů.....	- 59 -
<b>5.3</b>	<b>Databázový model.....</b>	<b>- 60 -</b>
5.3.1	ER-diagram .....	- 60 -

<b>6</b>	<b>ZHODNOCENÍ DOSAŽENÝCH VÝSLEDKŮ .....</b>	<b>- 61 -</b>
<b>6.1</b>	<b>Testování on-page SEO technik .....</b>	<b>- 61 -</b>
<b>6.2</b>	<b>Analýza odkazového profilu .....</b>	<b>- 62 -</b>
6.2.1	Porovnání výsledků s konkurenčními řešeními .....	- 63 -
6.2.2	Výsledky analýz .....	- 64 -
6.2.3	Výkonnost systému .....	- 65 -
<b>6.3</b>	<b>Budování zpětných odkazů .....</b>	<b>- 65 -</b>
6.3.1	Změna hodnocení webu .....	- 66 -
6.3.2	Změna pozic .....	- 66 -
<b>6.4</b>	<b>Výsledky stahování aukcí .....</b>	<b>- 68 -</b>
<b>7</b>	<b>NÁPADY NA VYLEPŠENÍ.....</b>	<b>- 69 -</b>
<b>8</b>	<b>ZÁVĚR .....</b>	<b>- 70 -</b>
	<b>LITERATURA .....</b>	<b>- 71 -</b>
	<b>PŘÍLOHY .....</b>	<b>- 73 -</b>
<b>A.</b>	<b>ADRESÁŘOVÁ STRUKTURA PŘILOŽENÉHO DVD.....</b>	<b>- 73 -</b>
<b>B.</b>	<b>DATOVÝ SLOVNÍK.....</b>	<b>- 74 -</b>

# 1 Úvod

Prostředí internetu se během posledních let razantním způsobem změnilo. Počet webových stránek a informací na nich obsažených se stále zvyšuje a změnil se také způsob, jakým je vytvářen webový obsah. Jeho tvorba již není výsadou pouze odborníků v této oblasti, ale může jej vytvářet téměř kdokoli, kdo má přístup k internetu. V minulosti samozřejmě fungovala různá diskuzní fóra, poté přišla éra osobních blogů, avšak největším průlomem v tomto směru se stal zrod sociálních sítí, který zapříčinil masové vytváření obsahu na internetu i bez technických znalostí. S tímto rostoucím počtem informací přibývá i práce vyhledávacích enginů, kteří se snaží servírovat co nejrelevantnější výsledky. Zejména společnost Google se nejvíce zasazuje o udržování kvality obsahu na internetu. Svými pravidly a doporučeními směřuje tvůrce webových stránek k vytváření přínosného a zajímavého obsahu. Weby, které tyto zásady nesplňují, mohou být značně znevýhodněny ve výsledcích vyhledávání. Proto vznikl obor zvaný SEO neboli optimalizace pro vyhledávače.

Diplomová práce se bude zabývat současnými trendy v optimalizaci pro vyhledávače. Popsány budou různé techniky a faktory, které pozitivně i negativně ovlivňují hodnocení webové stránky a její výsledné umístění. Dále budou zkoumány hodnotící algoritmy vyhledávačů Google a Seznam a jejich dopad na SEO trendy. Nejdůležitější částí SEO je odvětví zvané budování zpětných odkazů, jehož důležitost tkví v tom, že počet a kvalita zpětných odkazů udávají míru autority daného webu. Pro získání dobrého hodnocení, a tedy i dobrých pozic ve výsledcích vyhledávání, je však nutné při budování odkazů dodržovat určitá pravidla.

Text bude zaměřen hlavně na studium a testování různých způsobů tvorby a získávání nových odkazů. Důležitou součástí práce bude i analýza existujících nástrojů řešících tuto problematiku. U vybraných řešení budou popsány jejich funkcionality, tarify, výhody a nevýhody.

Hlavním cílem diplomové práce bude implementace vlastního nástroje sloužícího ke komplexní analýze odkazového profilu webu a budování zpětných odkazů, a proto jí bude věnována největší část textu. Funkcionalita systému bude založena na poznatcích získaných z úvodních kapitol. Bude uveden popis jednotlivých funkcí a průběhu implementace, včetně procesu optimalizace. Součástí této kapitoly bude i popsání tvorby modulu pro stahování aukcí z portálu Aukro.

V závěru práce zhodnotím data získaná testováním aplikace na reálných projektech a provedu jejich srovnání s konkurenčními řešeními. Zaměřím se jak na výsledky analýz odkazových profilů, tak i na dopad na hodnocení webu v průběhu budování nových odkazů.

## 2 Search Engine Optimization (SEO)

### 2.1 Co je to SEO?

Jedná se o optimalizaci webových stránek pro vyhledávače zahrnující soubor konkrétních technik a postupů, které mají vést k lepšímu umístění stránek ve výsledcích vyhledávání. Proces by měl být složen ze čtyř základních částí prováděných iterativně, dokud nedojde k plánovanému výsledku.

1. Analýza současného stavu webu za použití nástrojů (měření pozic, analýza klíčových slov, copywriting, zpětné odkazy aj.)
2. Návrh změn a postupů
3. Aplikování navržených postupů
4. Měření výsledků

Nové trendy v oblasti SEO jsou z největší části udávány společností Google. Jejím hlavním cílem je obrátit smysl této disciplíny od optimalizace pro vyhledávače spíše k optimalizaci pro návštěvníky stránek. V praxi to například znamená, aby obsah internetové stránky odpovídal očekávání uživatele. Dalším cílem je snaha vymýtit umístování zpětných odkazů na link farmy kvůli zvýšení hodnocení webu. Google vyvíjí stále lepší algoritmy, které tyto nekalé praktiky odhalují a snižují interesovaným webům jejich rank. V poslední době se stále více hovoří o tom, že SEO je „mrtvé“, ale podle mého názoru se stalo pouze složitějším a je potřeba mu věnovat více času.

### 2.2 SEO faktory

Existuje velké množství faktorů, které ovlivňují kvalitu a hodnocení internetových stránek. Ovlivnění může být pozitivní, ale i negativní. Každý vyhledávací engine hodnotí stránky různými metrikami, které nejsou zcela známy, protože si je jednotlivé společnosti pečlivě střeží. Navenek prezentují pouze výčet nejlepších SEO praktik. Jsou však prováděny různé studie a testování, které alespoň částečně odhalují důležitost jednotlivých faktorů, jež se mohou vyskytovat buď přímo na webu (on-page), nebo mohou působit z vnějšku (off-page). Velmi často se praktikují techniky, které mají na úkor kvality obsahu obelstít vyhledávací roboty za účelem získání lepší pozice v SERP. Těmto technikám se říká černé SEO a jsou v současnosti odhalovány zejména společností Google, která na ně reaguje penalizací a snížením pozic.

#### 2.2.1 On-page

##### **Umístění klíčových slov**

Při optimalizaci webové stránky na určité klíčové slovo nebo frázi je nutné jej vhodně umístit do různých jejích částí. Nejdůležitějším místem z pohledu vyhledávacích engineů je meta titulek stránky. V textu by měla být klíčová fráze rozmístěna rovnoměrně tak, aby obsah působil

přirozeně a nejednalo se pouze o výčet slov. Dále je vhodné používat slova v nadpisech H1-H6, URL adrese, meta popisu, alternativním textu obrázků, v textech interních i externích odkazů nebo v tučně zvýrazněném textu. Uvádí se, že hlavní klíčové slovo by se mělo vyskytovat v prvních 50-100 slovech HTML textu.

## **Hustota klíčových slov**

Jedním z hodnotících faktorů je také hustota klíčových slov v obsahu stránky. Je to podíl počtu klíčových slov vzhledem k celkovému počtu slov. Optimální hodnota se uvádí mezi 5-7 %. Přesné číslo však vyhledávací společnosti nesdělují. Při menší hustotě se může stát, že robot vyhodnotí stránku jako méně optimalizovanou na dané klíčové slovo a v opačném případě, tedy při jejím překročení, si může myslet, že jde o tzv. *keyword stuffing* neboli nepřiměřeně velké použití klíčových slov. Oba tyto případy mohou vést k penalizaci ve výsledcích vyhledávání. Pro měření hustoty existuje mnoho online nástrojů jako například <http://seo-servis.cz/keywords-test-klicovych-slov> nebo <http://www.keyworddensity.com>.

## **URL adresa**

Optimalizace URL adresy stránky neslouží pouze pro vyhledávací roboty, ale hlavně pro její uživatele. Z pohledu vyhledávacích enginů je žádoucí, aby obsahovala klíčová slova relevantní k obsahu stránky. Uživateli by měla podávat informaci o tom, co se na stránce nachází, být snadno zapamatovatelná a v co nejkratším tvaru. Pro lepší pochopení kontextu je vhodné uvádět v URL adrese například kategorie a podkategorie, pokud je to možné. Ze studií společnosti SEOMoz vyplývá, že kratší URL adresy s klíčovými slovy umístěnými blíže k názvu domény, mají tendenci být lépe hodnoceny. Samozřejmostí je odstranění přípony souboru, které nejen zkrátí adresu, ale zároveň zajistí univerzálnost platformy. [7]

Příklad optimalizované URL - [www.priklad.cz/kategorie/podkategorie/nazev-produktu](http://www.priklad.cz/kategorie/podkategorie/nazev-produktu)

## **Meta značka TITLE**

Jedná se o jeden z nejdůležitějších SEO faktorů na stránce. Aby se stránky zobrazovaly ve výsledcích vyhledávání na konkrétní klíčová slova nebo fráze, je potřeba je umístit právě do tohoto elementu. Správně optimalizovaný titulek má pozitivní vliv nejenom na hodnocení vyhledávacích enginů, ale také může zajistit vyšší míru prokliku od samotných uživatelů. Jeho hlavním cílem je ukázat, o čem stránka pojednává. Měl by tedy obsahovat primární a sekundární klíčová slova, případně obchodní značku a neměl by mít více než 70 znaků. Vyhledávací enginy totiž zkrátí v SERP obsah titulků zhruba na 65 znaků a zbytek se nahradí „...“ [7]. Je také známo, že vyhledávače hodnotí unikátnost titulků v rámci jednoho webu, a proto by každá stránka měla mít svůj vlastní. Titulky stránek se také často používají jako text odkazů na sociálních sítích, a proto přispívají k relevanci daného webu při budování zpětných odkazů.

## **Meta značka DESCRIPTION**

Meta popis neslouží přímo ke zvýšení ranku z pohledu vyhledávače, ale má sloužit spíše jako bezplatná reklama, která má uživatele upoutat a přimět k tomu, aby navštívil právě danou

stránku. Je vhodným místem pro použití klíčových slov a frází vzhledem k jejich tučnému zobrazení v SERP při shodě s vyhledávacím dotazem. Správně použitý popis může značně zvýšit míru prokliku. Naopak vytvoření špatného popisu může uživatele odradit, i když bude web na vysoké pozici v SERP.

## **Meta značka KEYWORDS**

Meta značka „keywords“ byla v minulosti hojně využívána některými vyhledávacími roboty pro určení pozice ve výsledcích vyhledávání. Proto čím dál častěji docházelo k jejímu zneužívání a tím také manipulaci s výsledky. Z tohoto důvodu nemá v současnosti téměř žádný význam při hodnocení ranku a pozice v SERP. V určitém směru může být i škodlivá. Například pokud je web používán pro komerční účely, mohou být klíčová slova obsažená ve značce „keywords“ vykradena konkurencí.

## **Obrázky**

Moderní weby se již bez obrázků téměř neobejdou, a proto se staly jedním z faktorů při jejich hodnocení. Vyhledávací enginy zatím nejsou schopny číst vizuální obsah obrázků a je tedy potřeba jim ho sdělit jiným způsobem. Doporučuje se používat textový atribut „alt“, který se zobrazuje v případě, že obrázek není možné vykreslit. Měl by obsahovat popis obrázku zahrnující klíčová slova relevantní k obsahu webové stránky. Je vhodné také doplnit text stejného významu před a za obrázek, aby vyhledávací robot získal potřebný kontext. Pokud je obrázek zároveň i odkazem na jinou stránku, může mít stejnou důležitost jako klasický odkaz a je tedy vhodné alternativní text obohatit o relevantní klíčové výrazy.

## **Atribut rel=“nofollow“**

Odkaz opatřený tímto atributem sloužil dříve k tomu, aby vyhledávači sdělil, že jej má ignorovat, a proto se s ním při výpočtu hodnoty odkazu nepočítalo. Společnost Google však v roce 2009 vydala prohlášení, že změnila jeho zpracování. Nyní se s odkazem pracuje stejně jako s ostatními, ale s tím rozdílem, že nepředává cílové stránce žádnou odkazovou popularitu. Tedy pokud na stránce existuje 10 odkazů a z toho jsou 2 s atributem nofollow, bude mít každý odkaz hodnotu  $X/10$  (dříve by to bylo  $X/8$ ). Tyto odkazy je vhodné používat zejména pro uživatelské komentáře kvůli omezení spamu.

## **Atribut rel=“canonical“**

Tento atribut sděluje vyhledávači, jakou stránku s duplicitním obsahem má považovat za hlavní. Používá se v situaci, kdy zobrazujeme podobný obsah a mění se pouze parametry URL adresy například při filtrování. Kdyby engine indexoval desítky až stovky URL adres s duplicitním obsahem, snížil by hodnocení celého webu. V této situaci je vhodné uvést jeden hlavní web, který se má zobrazit v SERP. Atribut se vkládá do sekce <head> na všechny stránky s duplicitním obsahem.



## **Javascript a FLASH**

Javascript a FLASH jsou technologie, které sice vizuálně působí lépe než obyčejné HTML, ale pro vyhledávací roboty je velmi těžké zjistit jejich sémantiku. Přestože vyhledávací enginy jsou schopny přečíst jejich zdrojový kód a stále pracují na vylepšení algoritmů k jejich analýze, není zaručeno správné pochopení významu, což může vést ke snížení ranku. Tyto grafické prvky by se neměly používat na úkor klasického obsahu, pokud nejsou podpořeny jinou technologií na jejich interpretování v HTML, například Swobject [7].

## **Přesměrování**

K přesměrování URL adresy na jinou URL adresu se používá mnoho způsobů například 301, 302, javascript nebo refresh atribut značky <meta>. Při přesměrování se však ztrácí hodnota odkazu, a proto je nutné vybrat správnou metodu. Společnost SEOMoz uvádí, že nejlepším typem je 301 redirect, který přenáší 90-99 % hodnoty odkazu. I přes svou ztrátovost je nesrovnatelně lepší než 302 redirect, který nepřenáší téměř žádnou hodnotu [7].

### **2.2.2 Off-page**

#### **Zpětné odkazy**

Jedná se o odkazy, které směřují na webovou stránku z cizích webů a mohou tím zvyšovat jeho autoritu. Zpětným odkazům se podrobněji věnuji v kapitole Budování zpětných odkazů.

#### **Stáří domény**

Webové stránky na starší doméně jsou obvykle považovány za autoritativnější než na nově vytvořené. Vyhledávač Google uvedl, že stáří domény je jedním hodnotících faktorů. Zkoumáno je datum prvního zaindexování obsahu a období, kdy na doménu začaly směřovat zpětné odkazy [8].

### **2.2.3 Black Hat SEO**

Jedná se o způsob „optimalizace“ stránek pomocí zakázaných technik. Někteří tvůrci webových stránek se uchylují k těmto nepovoleným praktikám, aby pro svůj web získali lepší rank nebo zpětné odkazy. Tyto techniky mohou nárazově pomoci zlepšit pozice ve výsledcích vyhledávání, ale jejich použití není z pohledu vyhledávacích enginů povoleno a při odhalení může dojít k penalizaci a výraznému zhoršení pozic v SERP.

#### **Cloaking**

Jedná se o techniku podstrčení jednoho obsahu vyhledávacímu robotovi a jiného obsahu návštěvníkovi. Kód nabídnutý vyhledávači je vysoce optimalizovaný na klíčové fráze a návštěvníkovi je naopak zobrazen uživatelsky přívětivý obsah.

## **Duplicita a kradení obsahu**

Jednou z nepovolených technik je i cílená duplicita obsahu na internetu. Vyskytuje se nejčastěji ve formě vykrádání cizích stránek a vkládání jejich textu na jiné stránky. Vyhledávače tuto praktiku velmi snadno odhalí a stránky nebo celý web může být penalizován. Při posuzování původu textu se přihlíží k datu indexace, stáří domény a autoritě odkazujících stránek. Dalším způsobem může být cílené budování více stránek se stejným obsahem s vidinou většího počtu stránek v SERP a tím větší pravděpodobnosti prokliku.

## **Skrytí textu**

Tato technika funguje tak, že je před uživatelem skryt určitý obsah na stránce, který může vidět jen vyhledávací robot. Docílit toho lze například splýváním barvy textu s pozadím, velmi malou velikostí písma, skrytím pomocí CSS nebo překrytím pomocí obrázku. Robotovi je předkládán optimalizovaný text, který by však uživateli nepřinesl žádnou hodnotu.

## **Doorway page**

Jde o způsob vysoké optimalizace vstupní stránky, která je zobrazována vyhledávačům s cílem zlepšit pozici webu. Většinou obsahuje až přehnaně velké množství klíčových slov. Samotná stránka však není zobrazována návštěvníkům, ale provádí pouze přesměrování na jinou stránku.

## **Link farmy**

Jsou to speciální weby, které obsahují stovky až tisíce stránek, které jsou vzájemně propojeny zpětnými odkazy. Stránkám na farmě se tak zvyšuje Page Rank, ale samy o sobě nemají žádnou informační hodnotu. Z pohledu vyhledávačů se jedná o jednu z nejhorších SEO technik a provozujícímu webu hrozí velká penalizace. Pokud odkazová farma směřuje na určitý web, nemělo by to na něj mít negativní vliv, protože by tato technika mohla být použita ke konkurenčnímu boji i bez vědomí majitele [8].

## **Klamavé přesměrování**

Když uživatel klikne na odkaz a zobrazí se mu jiný obsah, než očekával, jedná se o klamavé přesměrování. Nejčastěji se vyskytuje na stránkách s tematikou gamblerství, farmak apod. Je prováděno pomocí skriptů, které nejsou snadno odhalitelné vyhledávacím robotem, jako je například Javascript.

## **MFM**

„Made for money“ stránky se vyznačují tím, že samy o sobě nemají žádnou informační hodnotu a byly vytvořeny pouze pro vydělávání peněz z PPC reklam.

## 2.3 Budování zpětných odkazů

### 2.3.1 Důležitost zpětných odkazů

Zpětné odkazy jsou nejdůležitějším odvětvím optimalizace pro vyhledávače. Udává se, že SEO se skládá až z 70% právě z budování zpětných odkazů. Jistým způsobem se dá říct, že odkazy udávají míru oblíbenosti daného webu napříč internetem. Laicky řečeno internetové stránky, na které je hodně a kvalitně odkazováno, jsou vyhledávači hodnoceny lépe a dostávají se na vyšší pozice v SERP. Hodnocení je převzato ze způsobu uspořádání akademických prací. Jejich kvalita se určovala podle toho, kolik jiných akademických prací ji citovalo. Čím více je dokument (stránka) kvalitně odkazován, tím je hodnocen lépe a zvyšuje se jeho autorita. V současnosti nerozhoduje pouze celkový počet odkazů, ale v první řadě jejich kvalita. Odkazy z webů s vyšším hodnocením a z autoritativních webů přinášejí více „link juice“ neboli hodnoty odkazu.

Soubor všech zpětných odkazů webu se nazývá odkazový profil. Při jeho vyhodnocování se přihlíží k mnoha faktorům, jako jsou například text odkazu, link value, hodnocení a autorita odkazujícího webu, relevance obsahu, ale také jeho umístění odkazu na stránce. Důležitým ukazatelem kvality je i nárůst odkazů v čase. Přírůstek by měl být časově rovnoměrně rozložený, jinak by nepůsobil přirozeně. Počet odkazů by měl vykazovat trvalý nárůst, a proto nejsou vhodné nejružnější dočasné linky na nekvalitních serverech. Důležitá je i různorodost textů odkazů, typů odkazujících stránek a cílových stránek webu.

### 2.3.2 Techniky budování zpětných odkazů

Zpětné odkazy by neměly sloužit pouze k vylepšení hodnocení webu, ale hlavně k zajištění návštěvnosti. Mnoho lidí se neprávne domnívá, že budování odkazů slouží pouze pro uspokojení vyhledávačů, což není úplně pravda. Hlavním cílem by mělo být zejména přivedení nových zákazníků na webovou stránku. Největší idea je vytvořit přínosný a zajímavý obsah tak, aby na něj lidé odkazovali sami. To však není úplně jednoduché a ne každému se to podaří. Proto je téměř nezbytné vylepšovat svůj odkazový profil i ručně. Pro dosažení kýžených výsledků je potřeba dodržovat mnoho zásad a technik, díky nimž budou odkazy kvalitní a přinesou požadovanou hodnotu. Nepromyšlené a masové budování zpětných odkazů může být ve výsledku zcela zbytečné a v horším případě může mít i negativní vliv. Je lepší mít méně, ale zato kvalitních zpětných odkazů, než větší počet nekvalitních.

#### **Text odkazu**

Text odkazu býval velmi důležitý při určování pozic na různá klíčová slova, ale postupem času se jeho váha snížila. Příčinou byly tzv. „google bomby“. Uživatelé odkazovali na určitý web stejným slovním spojením a ten se poté začal objevovat ve výsledcích vyhledávání [8].

Při budování odkazů by měl být jejich text vhodně optimalizován na klíčové slovo, které je relevantní k cílové stránce, aby vyhledávací robot správně pochopil celkový kontext. Takový odkaz přinese stránce mnohem větší hodnotu, než kdyby byl složen pouze z názvu webu

nebo obecné fráze jako například „*www.domena.cz*“, „*Klikněte zde*“ nebo „*Odkaz*“. Na druhou stranu odkazový profil složený pouze z optimalizovaných textů by nevypadal přirozeně, a tudíž je vhodné se občas odkazovat i obecnými frázemi. Uživatelé totiž při odkazování SEO vůbec neřeší a většina z nich tento pojem ani nezná. Správný odkazový profil by měl tedy působit přirozeně a anchor texty by měly být rozmanité.

### **Link Value (Link Popularity)**

Hodnota nazývaná také jako link juice, kterou odkaz předává cílové stránce je nejdůležitější metrikou při budování zpětných odkazů. Závisí z největší části na hodnotě PR a na počtu odchozích odkazů na stránce, mezi které se dělí. Roli mohou hrát i faktory jako je relevance obsahu, stáří domény a jiné veličiny, které nejsou známy. U kvality odkazu se často hovoří i o autoritě webu, která je dána kvalitním odkazovým profilem, stářím domény, kvalitou obsahu a návštěvností. Při výběru vhodné odkazující stránky bychom měli věnovat pozornost především počtu odchozích odkazů a tematičnosti.

### **Relevance obsahu**

Link value je závislá i na obsahu odkazující stránky. Vyhledávače dávají větší přidanou hodnotu těm odkazům, které pocházejí ze stránek s podobným zaměřením. Představme si e-shop s podlahovými krytinami, na který odkazuje stránka o pokládce a údržbě podlah. Takový odkaz bude mnohokrát hodnotnější, než kdyby pocházel ze stránek mysliveckého sdružení.

### **Pozice odkazu**

Jedním z faktorů je i umístění zpětného odkazu na stránce. Doporučuje se obklopit odkaz relevantním textem, to znamená takovým, který obsahuje podobná klíčová slova. Umístění odkazu do postranního bloku nebo patičky, který se opakuje na všech podstránkách, může zapříčinit snížení hodnoty všech odkazů z daného webu. Vhodné je vytvořit speciální stránku pro účely odkazování nebo odkazovat přímo z hlavního obsahu.

## **2.3.3 Definování oblastí pro budování zpětných odkazů**

### **Katalogy**

Nejznámějším způsobem budování odkazů jsou internetové katalogy. V současnosti jim však vyhledávače nepřikládají tak velkou důležitost, jako tomu bylo v minulosti. Velké množství odkazů z nekvalitních katalogů může webu dokonce uškodit. Na druhou stranu jsou stále velmi oblíbenou a zatím nepostradatelnou součástí link-buildingu. Samozřejmostí je registrace u známých katalogů jako jsou firmy.cz, centrum.cz aj. Obecně platí, že čím méně je odkazů na jedné straně katalogu, tím více to přináší hodnoty odkazovaným stránkám. Před zaregistrováním je tedy vhodné si katalog prohlédnout a případně zkontrolovat potřebné metriky. Pro web jsou prospěšnější ty katalogy, které se zabývají stejným nebo podobným odvětvím. Vyhledávače tento odkaz hodnotí velmi kladně. Kombinací správného počtu externích odkazů na stránce, dobré hodnoty Page Ranku a relevantním obsahem se odkaz stává kvalitním a může zlepšit výsledný rank webu. V opačném případě může mít odkaz i negativní vliv na pozici v SERP.

## **PR články**

Tento typ článků můžeme chápat jako reklamní text, který má přilákat návštěvníky na daný web. Články se většinou skládají z titulku a textu okolo 250 – 500 znaků. Při jeho psaní je důležitá jeho optimalizace na klíčové slovo nebo frázi tak, aby se vyskytovalo jak v titulku, tak i se správnou hustotou v textu. Je potřeba se odkazovat na stránku, která je relevantní k psanému textu. Jako u veškerého obsahu na internetu je i zde nutné si dát pozor na duplicitu. Při tvorbě PR článků je tedy vhodné jejich obsah obměňovat. Existuje velké množství serverů, které poskytují publikování těchto článků jak zdarma, tak i v placené podobě.

## **Sociální sítě**

Snaha vyhledávacích enginů o zkvalitnění obsahu na internetu přinesla v posledních letech změny v hodnocení zpětných odkazů. Weby, jejichž odkazový profil je přirozený a s pozvolným nárůstem jsou vnímány jako kvalitnější a jejich pozice jsou zvýhodňovány. Tento trend podporují i sociální sítě, kde je obsah z větší části tvořen samotnými lidmi a ne pouze pro účely SEO. Tudíž i odkazy jsou tvořeny přirozenou cestou a jejich počet se odvíjí podle toho, jak je webová stránka oblíbená nebo jak zajímavý obsah nabízí. V této oblasti je tedy cílem vytvořit tak zajímavý obsah, aby na něj lidé sami odkazovali a sdíleli jej s přáteli. Google již několikrát prohlásoval, že aktivita a odkazování webu na sítích Facebook a Twitter a Google + má vliv na jeho hodnocení. Je tedy velmi praktické umísťovat na stránku ikony umožňující lidem snadné odkazování a sdílení obsahu.

## **Podpůrné weby**

Jsou to menší weby (microsite, minisite), které slouží k podpoře hlavního webového projektu. Vytvořením poutavého a zajímavého obsahu, na který bude mnoho lidí odkazovat z různých oblastí, získáme web s dobrým hodnocením, který bude zároveň relevantní k našemu hlavnímu webu. Poté stačí vhodně směřovat zpětné odkazy na jednu či více stránek kořenového webu a dostaneme tím kýžený „link juice“. Existuje mnoho způsobů jak tuto techniku využít. Představme si například e-shop nabízející kolečkové brusle, ke kterému vytvoříme podpůrný web poskytující informace o technikách bruslení, údržbě, servisu, výběru bruslí atp. Na něj umístíme odkazy na e-shop a pokud bude obsah přínosný a zajímavý natolik, že se na něj budou lidé odkazovat, získáme tím hodnotné zpětné odkazy. Takto vytvořená stránka navíc může přinést nové návštěvníky. Aby podpůrná stránka působila přirozeně, neměla by odkazovat pouze na jeden web a také se doporučuje ji umístit na jiný server než je odkazovaný web.

## **Oborové weby**

Jak již bylo zmíněno výše, jedním z nejdůležitějších metrik je relevance obsahu. Zpětné odkazy umístěné na webech se stejnou tematikou mohou přinést více „link juice“. Ideální je vyhledat a oslovit majitele webových stránek, které se věnují stejnému odvětví (nemusí se vždy jednat o konkurenci) a vložit zde svůj odkaz. V této oblasti link-buildingu se jedná nejčastěji o oboustrannou výměnu odkazů.

## **Blogy**

Dalším způsobem jak rozšířit svůj odkazový profil je vytvoření článku na blogu. Jsou tři možnosti jak toho docílit. První je vytvoření vlastního blogu, což není špatné řešení, ale takto založený web nebude mít potřebnou autoritu a návštěvnost. Na druhou stranu, pokud bude obsahovat zajímavý článek, který lidi zaujme a bude odkazován, může přinést jak návštěvníky, tak i kvalitní odkazy. Druhou možností je blogování na známých portálech, které mají svou historii a tudíž i vyšší návštěvnost a vyhledávače je často navštěvují. Pokud bude článek dost atraktivní a dostane se například na hlavní stranu, může správně použitý zpětný odkaz na vlastní web přinést větší hodnotu. Třetí variantou je tzv. guestblogging neboli hostující blogování. Princip spočívá v tom, že je vlastní článek nabídnut cizímu blogu k publikování. Po nalezení vhodného webu a oslovení majitele se nedoporučuje odkazovat hned v prvním článku na svůj web, ale budovat si postupně svoji pověst [11]. Tato oblast budování není pro každého a je nutné mít fantazii a smysl pro psaní nebo si najmout zkušeného copywritera, který dokáže napsat kvalitní a zajímavý článek.

## **Diskuzní fóra**

Komentování a odkazování na své stránky v diskuzních fórech neslouží ani tak k budování hodnotných odkazů, ale spíše k přivedení návštěvníků. Většina serverů obvykle cíleně nastavuje všem odkazům v komentářích atribut `rel="nofollow"`, aby nemohly přenášet žádnou hodnotu a diskuze se nepřeměnila ve spamování zpětnými odkazy. Pokud je však příspěvek obsahově kvalitní a relevantní, může čtenáře upoutat a ten poté navštíví danou stránku přes uvedený odkaz. Často se v diskuzích pod příspěvkem objevuje i tzv. podpis, což je také jedna z variant jak zobrazit svou URL adresu. Pokud je uživatel na diskuzním fóru zaregistrován, všechny jeho příspěvky odkazují na jeho profil, kde může mít uveden zpětný odkaz na svůj web a poté každým kvalitním příspěvkem zvyšuje jeho hodnotu [11].

## **Kupování odkazů**

Tato forma link-buildingu patří k těm nejsnadnějším. Za určitý obnos se jednoduše a bez námahy dá vytvořit bohatý odkazový profil, který však tvůrci vyhledávačů nemají rádi. Pokud se nejedná o link farmy, je těžké odhalit, zda je odkaz přirozený nebo byl zakoupen. Mnoho serverů na prodej odkazů provádí jejich ruční vkládání na různá místa webu podle přání klienta, a pokud se navíc jedná o tematicky podobnou stránku, neměl by to vyhledávač odhalit. Škodlivé mohou být služby, které odkazy vkládají automaticky, a tudíž není dopředu jasné kde, a v jaké podobě se opravdu objeví [11]. Nákup odkazů sice není doporučovanou metodou, ale dobrým pojetím lze touto investicí získat kvalitní reference.

## Link Baiting

Jedná se o trochu odlišný způsob budování odkazů, který se nezabývá hledáním vhodných webů, ale vytvářením obsahu, který sám přiláká návštěvníky. Princip spočívá v tom, že na stránku s obsahem budou jeho čtenáři a uživatelé odkazovat sami, protože nabízí něco užitečného, zábavného, pobuřujícího nebo cokoliv zajímavého.

- **Online nástroje** – poskytnutí užitečného nástroje (výpočet PageRank, BMI, kurzy měn)
- **Návody** – lidé často rádi sdílejí a odkazují zajímavé návody, recepty, postupy (např. výměna a čištění koleček u in-line bruslí, cvičební plány apod.)
- **Soutěže, kvízy** – zajímavým způsobem jak přilákat návštěvníky je uspořádání soutěže, kvízu, ankety (nejlepší fotografie z dovolené, tipovací soutěže, vědomostní kvízy)
- **Zábava** – internetem se nejlépe šíří zábavný obsah; vytvořením nebo sdílením vtipného videa či obrázku na svém webu je možné získat mnoho přirozených návštěv a odkazů
- **Novinky, objevy** – lidi zajímají informace o aktuálním dění a jejich poskytování může zvýšit návštěvnost i odkazy
- **Kontroverze** – kontroverzní vystupování přináší vždy mnoho pozornosti a napsání zajímavého článku, který bude nekonvenční a navíc nabídne možnost vložit komentář, může přinést krátkodobý, ale zato velký příliv návštěv.

## „Broken link building“

Jde o poměrně novou techniku, která spočívá v hledání nefunkčních odkazů na autoritativních a tematicky podobných webech. Nalezený odkaz se poté nahlásí administrátorovi a jako alternativa se mu nabídne odkaz na vlastní web. Metoda má výhodu v tom, že žádný webmaster nechce mít na svých stránkách nefunkční odkazy a může tuto iniciativu ocenit. Na druhou stranu, vyhledání a oslovení je jen jedna část a poté záleží hlavně na vstřícnosti provozovatele. Nicméně se jedná o techniku, která může zajistit odkazy z ověřených a kvalitních webů, na které se za normálních okolností nedá dostat.

## 2.4 Vyhledávací enginy a jejich algoritmy

Před samotnou optimalizací webu je důležité znát, na co se vyhledávače zaměřují a jakým způsobem hodnotí jednotlivé faktory. Jinými slovy je potřeba znát, co vyhledávací enginy očekávají, že bude webová stránka splňovat. V této kapitole se budu věnovat popisu současných hodnotících algoritmů vyhledávačů a jejich důležitosti.

### 2.4.1 Google

Společnost Google je v současnosti předním udavatelem směru, co se týče vyhledávání obsahu na internetu. Svými algoritmy a pravidly se snaží, aby výsledky vyhledávání byly co nejrelevantnější k potřebám uživatelů. V posledních dvou letech došlo k výrazným změnám v algoritmech hodnotících webovou stránku. Kvůli těmto změnám bylo nutné přehodnotit celou filozofii optimalizace pro vyhledávače, která byla dlouho dobu více cílena spíše pro vyhledávací roboty

než pro samotné lidi. S tímto novým pohledem na věc se začal klást důraz na unikátnost a relevanci obsahu. Google se také postavil proti webům, které provádějí tzv. černé SEO a jiné nedovolené praktiky, což se po zavedení algoritmu Penguin 2.0 v červnu 2013 dramaticky promítlo v jejich výsledných pozicích. Změny nastaly i v budování zpětných odkazů. Snížila se účinnost registrací do katalogů a uměle vytvořených odkazů a zvýšil se význam odkazů ze sociálních sítí.

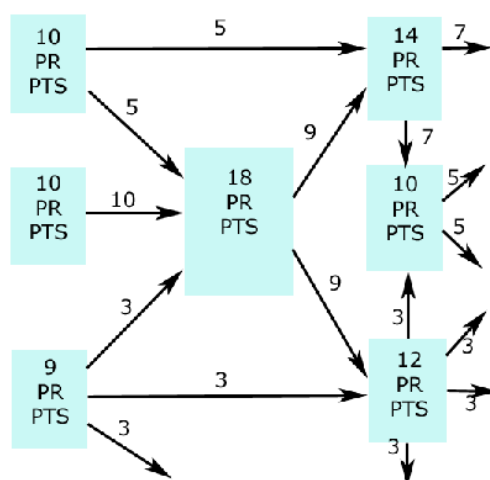
## Page Rank (PR)

Tento algoritmus hodnotí webové stránky podle odkazů vedoucích z jiných webů. Nezáleží jen na počtu zpětných odkazů, ale hlavně na jejich kvalitě. Jedná se o jedno z důležitých kritérií v oblasti SEO. Laicky řečeno se PR počítá podle počtu stránek odkazujících na danou stránku a jejich PR (Obrázek 1) [5]. Odkazující stránka T předává část svého page ranku  $PR(T)$ , která závisí na počtu odchozích odkazů  $C$ . Z toho vyplývá, že čím více odkazů stránka obsahuje, tím méně hodnoty předává.

$$PR(A) = (1 - d) + d (PR(T1)/C(T1) + \dots + PR(Tn)/C(Tn))$$

*Rovnice 1: Obecná rovnice pro výpočet Page Ranku*

Jedná se pouze o základní rovnici popsanou samotnými autory (Rovnice 1) [6]. Přesná definice hodnotícího algoritmu není známá a jistě obsahuje velké množství dalších proměnných.



*Obrázek 1: Graf obecné distribuce Page Ranku*

Do rovnice se vkládá i konstanta  $d$  (damping faktor), což je pravděpodobnost, že náhodný návštěvník zadá jinou stránku a přeruší tím řetězec odkazů. Tato hodnota je v intervalu  $<0;1>$  a nejčastěji se používá číslo 0,85 [6].



## Google Toolbar Page Rank (GTPR)

Hodnota PR, která je uživatelům poskytnuta pomocí Google Toolbar není přesnou reprezentací skutečného čísla, ale pouze jeho odvozením viz Tabulka 1 [5]. Stupnice GTPR může nabývat hodnot od 0 do 10 a je logaritmická, proto je získání každého dalšího stupně těžší než toho předchozího. Tato hodnota se nepoužívá k řazení výsledků a není moc často aktualizována. Uvádí se, že aktuální GTPR hodnota může být 60 až 90 dní stará [7].

Skutečný Page Rank	Google Toolbar Page Rank
0-9	0
10-99	1
100-999	2
1000-9999	3
10000-99999	4
100000-999999	5
1000000-9999999	6
10000000-99999999	7
100000000-999999999	8
1000000000-9999999999	9
10000000000-99999999999	10

Tabulka 1: Rozdíl mezi skutečným PR a GTPR

## Google Panda

Google Panda je algoritmus, který se již dlouhou dobu snaží o zkvalitnění internetového obsahu a poskytnutí relevantních výsledků při vyhledávání. Jeho poslední aktualizace v roce 2012 přinesla zásadní průlom v celém SEO odvětví a zejména v budování zpětných odkazů. Po jeho nasazení klesly průměrné pozice až o desítky procent. Zaměřuje se zejména na originalitu obsahu webu, kvalitní odkazy a relevanci. Hlavními sledovanými faktory jsou:

- Nízké procento originálního obsahu na webu
- Nekvalitní odkazy směřující na konkrétní stránku i celý web
- Návštěvníci tráví na webu krátkou dobu
- Významné procento duplicitního obsahu na webu
- Vysoká míra okamžitého opuštění webu (bounce rate)
- Významné procento obsahu stránky tvoří prvky, které se opakují na všech stránkách webu
- Významné množství nerelevantních reklamních prvků
- Nepřirozené používání klíčových slov na stránce

Pokud Panda odhalí některý z těchto prvků je pravděpodobné, že se to odrazí ve výsledných pozicích webu. Nejčastějším případem je masové psaní PR článků na různých serverech. Problémem není velké množství článků, pokud jsou všechny kvalitní a unikátní. Jakmile se však začne objevovat stejný text na více PR webech se zpětnými odkazy na určitou doménu, je to jasný signál nekvalitního link-buildingu a celý web je tímto algoritmem penalizován.

## Google Penguin

Jedná se v podstatě o „sesterský“ algoritmus Google Pandy, který má stejné poslání, ale přináší navíc i pár rozšíření. Penguin 2.0 vydaný 22. května 2013 provedl další změny, které citelně zasáhly do optimalizace všech webů a způsobily jejich penalizaci a propad v pozicích. Hlavní faktory, které se začaly důrazněji hodnotit, jsou následující:

- Přehnaná optimalizace webu
- Diverzifikace textů odkazů
- Spamování odkazy na diskuzních fórech
- Odkazy z nekvalitních webů
- Nákup odkazů
- Nerelevantní site-wide odkazy
- Odkazování pouze na jednu stránku domény

Je jasné vidět, že se algoritmus tvrdě postavil hlavně proti nekvalitnímu odkazovému profilu. Největší újmu utrpěly služby, které prodávají odkazy nebo nabízejí publikování článků. Na těchto webech se hojně objevovaly duplicitní obsahy s velkým množstvím odkazů, což způsobilo jejich penalizaci a některé byly dokonce úplně vyřazeny z indexu vyhledávače [3][4].

## Google Hummingbird

V polovině roku 2013 se objevil algoritmus Hummingbird, který byl největší aktualizací v Google za posledních 5 let. Nebyl již zaměřen na nedovolené SEO praktiky jako výše uvedené algoritmy, ale změnil koncept vyhledávání a zobrazení výsledků z pohledu samotného vyhledávacího enginu. Hlavním cílem bylo pochopení toho, co uživatel skutečně hledá a ne se odkázat pouze na klíčová slova v dotazu. Snažil se o poskytnutí hledaných informací již v SERP bez kliknutí na odkaz. Aktualizace zahrnovala:

- Konverzační hledání – výsledky závislé na předchozím hledání
- Snížení významu long-tail frází
- Personalizace výsledků

[2]

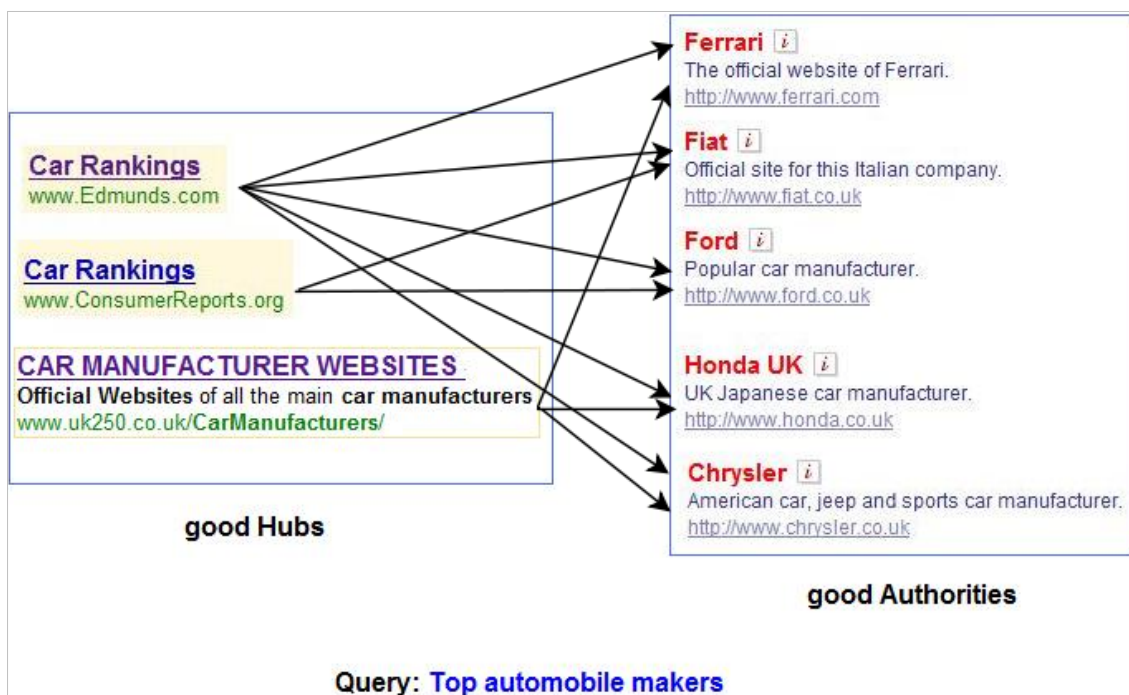
## 2.4.2 Seznam

### S-rank

Vyhledávač Seznam má svůj hodnotící algoritmus pro webové stránky s názvem S-rank, který se podobá Page Ranku. Jeho výpočet se odvíjí podle odkazového profilu webu a velkého množství dalších veličin. V potaz se bere nejen kvalita příchodích, ale i odchozích odkazů. Přesný postup výpočtu není veřejný, ale vychází z algoritmu Hubs & Authorities. Hodnota S-ranku, kterou je možno zjistit pomocí Seznam Lištičky může nabývat hodnot od 0 do 10. Stejně jako u Page Ranku se nejedná o skutečné číslo, ale pouze o jeho zaokrouhlení [1].

## Hubs & Authorities (HITS)

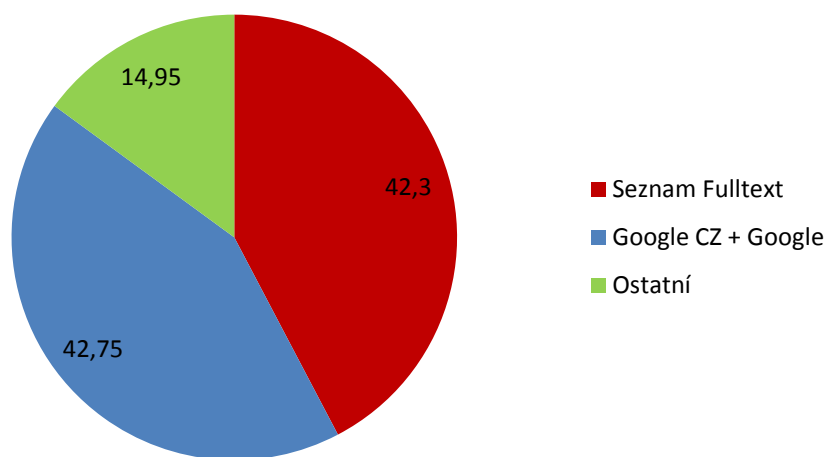
Jedná se o hodnotící algoritmus, který vznikl ve stejné době jako Page Rank. Hlavním kritériem je opět odkazový profil, ale stránka dostává dvojí hodnocení. Hodnota *autorita* udává informační kvalitu obsahu stránky, zatímco hodnota *hub* určuje kvalitu odchozích odkazů ze stránky. Jinými slovy Hub slouží k propagaci autoritativních stránek, od kterých uživatel očekává relevantní výsledek (Obrázek 2) [10]. Hodnota HITS se počítá vzájemnou rekurzí. *Autorita* se rovná sumě hodnot všech jejích *hubů*:  $\sum_{i=1}^n \mathbf{hub}(i)$ . *Hub* se počítá jako suma všech hodnot *autorit*, na které odkazuje:  $\sum_{i=1}^n \mathbf{auth}(i)$ . Z toho vyplývá, že dobrý *hub* zvyšuje *autoritu* stránky a naopak [10].



Obrázek 2: HITS algoritmus

### 2.4.3 Srovnání

Při srovnání kvality vyhledávání a řazení výsledků je bezpochyby nejlepším vyhledávačem Google. Ve světě je také tím nejpoužívanějším řešením. V ČR je však stále na první pozici vyhledávač Seznam. Tento trend je způsoben historickým aspektem, neboť mnoho lidí při začátku internetu neznalo jiné, než tuzemské řešení a Seznam se stal v Česku synonymem pro internet a vyhledávání. Velkou zásluhu na tom měly i přidružené služby jako e-mailová schránka, *firmy.cz*, *lidé.cz* aj., díky kterým jej většina uživatelů považovala za centrum všeho dění na internetu a mnohým to zůstalo dodnes. Rok od roku se náskok Seznamu před Googlem snižuje a v brzké době se očekává jeho smazání. Podle webu *TOPList.cz* je v současnosti jejich používání vyrovnané (Obrázek 3) [12].



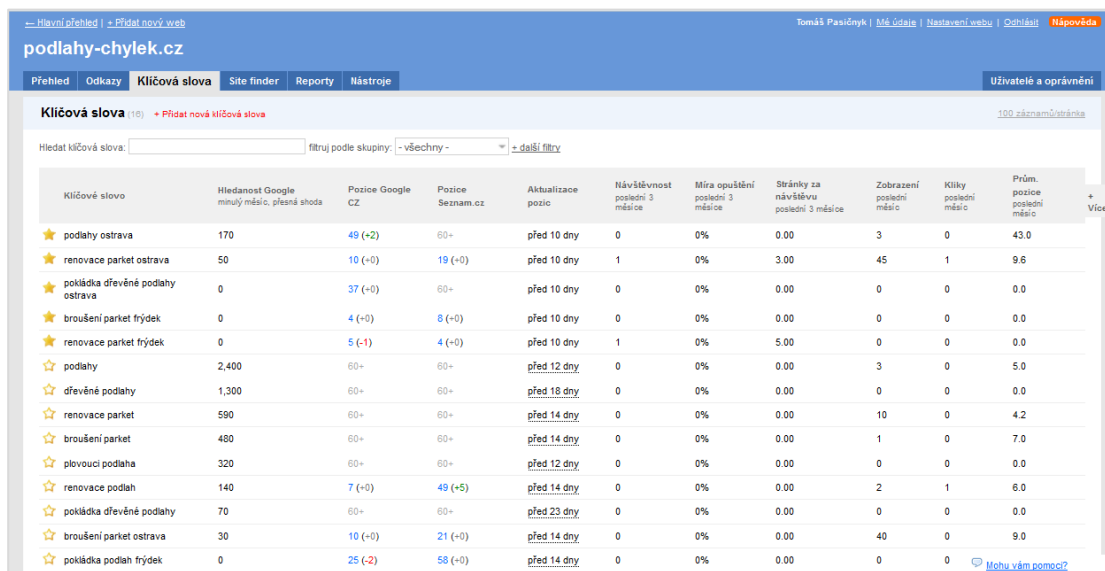
Obrázek 3: Graf použití vyhledávačů v ČR - týden 15/2014

## 3 Existující řešení pro SEO analýzu

V současnosti existuje nepřeberné množství různých nástrojů, které se zabývají SEO analýzou a hledáním zpětných odkazů. Před vývojem vlastního nástroje jsem prozkoumal a otestoval několik konkurenčních řešení, abych zjistil jejich funkcionality a mohl si jimi inspirovat. Vybral jsem jak z českých, tak i zahraničních nástrojů. Každý z nich poskytuje komplexní nabídku funkcí, avšak při jejich popisu jsem se zaměřil hlavně na ty, které se vztahují k budování zpětných odkazů.

### 3.1 Collabim

Collabim je český nástroj pro SEO analýzu, který je zaměřen spíše na zjištění pozic klíčových slov ve vyhledávačích. Nabízí však i funkci Site Finder, která slouží k vyhledání podobných webů pro výměnu odkazů. Poskytuje také možnost kontroly zpětných odkazů. Projekt je možné propojit s účtem Google Analytics a Google Webmaster Tools pro sledování návštěvnosti a konverze.



The screenshot shows the Collabim web interface. At the top, there's a navigation bar with links like 'Hlavní přehled', 'Přidat nový web', and user settings. Below this, a breadcrumb trail shows 'podlahy-chylek.cz' > 'Klíčová slova'. The main section is titled 'Klíčová slova' and contains a search bar and a table of keyword data. The table has columns for keyword, Google search volume, Google position, Seznam.cz position, update date, last 3 months traffic, last 3 months bounce rate, last 3 months pages, last month impressions, last month clicks, and average position. The keywords listed include 'podlahy ostrava', 'renovace parket ostrava', 'pokládka dřevěné podlahy ostrava', etc.

Klíčové slovo	Hledanost Google minulý měsíc, přesná shoda	Pozice Google CZ	Pozice Seznam.cz	Aktualizace pozic	Návštěvnost poslední 3 měsíce	Míra opuštění poslední 3 měsíce	Stránky za návštěvu poslední 3 měsíce	Zobrazení poslední měsíc	Kliky poslední měsíc	Prům. pozice poslední měsíc	+ Více
★ podlahy ostrava	170	49 (+2)	60+	před 10 dny	0	0%	0.00	3	0	43.0	
★ renovace parket ostrava	50	10 (+0)	19 (+0)	před 10 dny	1	0%	3.00	45	1	9.6	
★ pokládka dřevěné podlahy ostrava	0	37 (+0)	60+	před 10 dny	0	0%	0.00	0	0	0.0	
★ broušení parket frýdek	0	4 (+0)	8 (+0)	před 10 dny	0	0%	0.00	0	0	0.0	
★ renovace parket frýdek	0	5 (-1)	4 (-0)	před 10 dny	1	0%	5.00	0	0	0.0	
☆ podlahy	2,400	60+	60+	před 12 dny	0	0%	0.00	3	0	5.0	
☆ dřevěné podlahy	1,300	60+	60+	před 18 dny	0	0%	0.00	0	0	0.0	
☆ renovace parket	590	60+	60+	před 14 dny	0	0%	0.00	10	0	4.2	
☆ broušení parket	480	60+	60+	před 14 dny	0	0%	0.00	1	0	7.0	
☆ plovoucí podlaha	320	60+	60+	před 12 dny	0	0%	0.00	0	0	0.0	
☆ renovace podlah	140	7 (+0)	49 (+5)	před 14 dny	0	0%	0.00	2	1	6.0	
☆ pokládka dřevěné podlahy	70	60+	60+	před 23 dny	0	0%	0.00	0	0	0.0	
☆ broušení parket ostrava	30	10 (+0)	21 (+0)	před 14 dny	0	0%	0.00	40	0	9.0	
☆ pokládka podlah frýdek	0	25 (-2)	58 (+0)	před 14 dny	0	0%	0.00	0	0		Mohu vám pomoci?

Obrázek 4: Ukázka nástroje Collabim

#### 3.1.1 Funkce

##### Měření pozic ve vyhledávačích

Umožňuje sledovat pozice až 1500 klíčových slov ve vyhledávačích Google a Seznam. Vedle pozic je zobrazena i hledanost fráze na Google za minulý měsíc. Přidáním přihlašovacích údajů do Google Analytics získáme návštěvnost, míru opuštění, konverzní poměr aj. Dále je možné u klíčového slova sledovat, v kolika získaných zpětných odkazech se vyskytuje (Obrázek 4). Měření pozic probíhá pomocí rozšíření v prohlížečích uživatelů, bez jehož instalace je

aktualizace funkční pouze 30 dní. Pozice pro vybraná klíčová slova jsou aktualizovány jednou denně (v neplacené verzi omezeno na 5 frází), zbytek jednou za 14 dní. Do projektu je možné přidat více sledovaných webů, například konkurenčních. Omezení jejich počtu závisí na druhu tarifu.

## Site Finder

Slouží k vyhledání příbuzných webů s podobným obsahem, které by mohly být vhodné pro vložení zpětného odkazu. Weby nachází podle předem vložených klíčových slov a frází. Umožňuje filtrování na komerční a nekomerční weby, diskuzní fóra a blogy. U každého navrženého webu je zobrazen počet jeho zpětných odkazů, počet externích odkazů a počet stránek.

## Zpětné odkazy

Poskytuje nástroj na správu a kontrolu zpětných odkazů. Neslouží k jejich vyhledávání, ale pouze k evidenci a kontrole. Odkazy se přidávají ručně nebo hromadným importem a uvádí se jejich:

- **typ** – výměnný odkaz, registrace do katalogu, článek, banner, diskuzní příspěvek nebo neurčitý
- **stav** – čeká, kontaktováno, dokončeno
- **URL** adresa odkazující stránky
- datum aktivity od-do (nepovinné)

Systém poté načte informace o odkazu, kterými jsou text, cílová stránka a výsledek kontroly (Obrázek 5). Funkcionalita hlídá jejich existenci na odkazujících webech a je užitečná zejména pro výměnné nebo placené odkazy. U placených odkazů je možnost evidence jejich ceny, plateb a faktur.

Typ	Výsledek kontroly odkazů	URL stránky s mým odkazem	Stav	Je aktivní?	Den přidání	Text odkazu	Cílová stránka
Odkaz	Odkaz ještě nebyl zkontrolován	<a href="http://www.atlasfrem.info/katalog/bydle...">http://www.atlasfrem.info/katalog/bydle...</a>	dokončeno	neomezeno	16.4.2014	<a href="http://www.podlahy-chylek.cz">www.podlahy-chylek.cz</a>	Podlahy Martin Chýlek   Moravskoslezský kraj, Frýdek-Místek, Frydlant nad Ostravicí, Ostrava, Frenštát pod Radhoštěm, Rožnov pod Radhoštěm - Úvod
Odkaz	Odkaz ještě nebyl zkontrolován	<a href="http://www.kontaktyfrem.cz/detail/marti...">http://www.kontaktyfrem.cz/detail/marti...</a>	dokončeno	neomezeno	16.4.2014	<a href="http://www.podlahy-chylek.cz">http://www.podlahy-chylek.cz</a>	Podlahy Martin Chýlek   Moravskoslezský kraj, Frýdek-Místek, Frydlant nad Ostravicí, Ostrava, Frenštát pod Radhoštěm, Rožnov pod Radhoštěm - Úvod
Odkaz	Odkaz ještě nebyl zkontrolován	<a href="http://www.malovani-stein.cz/">http://www.malovani-stein.cz/</a>	dokončeno	neomezeno	16.4.2014	-	-

Obrázek 5: Ukázka evidence odkazů nástroje Collabim

## Import

Pro snadnější nahrávání informací o projektech je k dispozici import webů, klíčových slov a odkazů pomocí souborů typu xls a xlsx. Ke každému případu je poskytován vzorový dokument.

## Reporty

Nástroj nabízí generování pravidelných reportů jednotlivých analýz a jejich zasílání na e-mailovou adresu ve formátech ODT a PDF, které obsahují:

- Tabulku s aktuálními pozicemi vybraných nebrandových i brandových klíčových slov ve vyhledávačích Google CZ a Seznam
- Grafy s vývojem pozic u jednotlivých klíčových slov v obou vyhledávačích
- Seznam získaných zpětných odkazů v daném časovém období
- Návštěvnost a konverze z Google Analytics

### 3.1.2 Ceník tarifů

Ceny tarifů se odvíjí zejména podle počtu zkoumaných webů (Tabulka 2). Ve verzi zdarma je omezena pouze funkce Site Finder, která funguje jen 30 dní a možnost zasílání reportů, která není poskytována vůbec. Omezen je i počet klíčových slov, která jsou aktualizována denně.

	FREE	Basic	Standard	Plus	Premium
<b>Počet webů</b>	1	5	15	40	100
<b>Klíčová slova</b>	5	40	40	40	40
<b>Site Finder</b>	NE	ANO	ANO	ANO	ANO
<b>Reporty</b>	NE	ANO	ANO	ANO	ANO
<b>Cena (měsíc)</b>	0	450	1200	3000	6000

Tabulka 2: Ceník tarifů Collabim

## 3.2 SeoNástroje

Internetový projekt SEO Nástroje je dalším z mnoha českých SEO analyzátorů. Ke srovnání jsem ho vybral z důvodu, že nabízí více, než jen audit webu a zjištění pozic u klíčových slov. Poskytuje navíc srovnání dvou webů, registraci do katalogů a tvorbu microsites.

The screenshot shows the SEO Nástroje website interface. At the top, there's a navigation bar with links like 'Úvod', 'Co je cílem SEO?', 'Jak Vám SEO nástroje pomohou?', and 'Kontakt'. Below this, the main content area displays an 'SEO analýza' report for the URL 'http://www.podlahy-chylek.cz' with a score of 79%. The report includes a table of 'Základní informace' (Basic Information) with attributes like Content-type, Robots, and Keywords. The left sidebar contains a menu for 'SEO nástroje (zdarma)' (Free SEO tools) and 'Novinky' (News). The right sidebar lists 'SEO služby (placené)' (Paid SEO services) and 'Další užitečné projekty' (Other useful projects).

Atribut	Hodnota
Content-type	text/html, charset=utf-8
Robots	index, follow
Keywords	podlahy, podlaharské práce, podlaharství, martin chylek, chylek, dřevěné podlahy, parkety, renovace podlah, pvc, laminátové podlahy, vyrovnávání podlah, renovace schodů, koberec, fytex, misky, fytex, renovace podlah, renovace schodů, koberec, fytex, misky, fytex, renovace podlah, renovace schodů, koberec, fytex, misky, fytex

Obrázek 6: Ukázka nástroje SEOnástroje.cz

### SEO analýza

Po zadání URL adresy zkoumaného webu provede audit on-page SEO faktorů a vypíše výsledné skóre od 0 do 100 % (Obrázek 6):

- Znaková sada
- Titulek stránky
- Popis stránky
- Robots.txt
- Sitemap.xml
- Velikost v kB
- Strukturu nadpisů
- Strukturu stránky
- Inline CSS styly
- Počet odkazů
- Počet externích odkazů
- Vnořené tabulky
- Nesémantické elementy
- Alternativní obsah
- Validitu stránky



## **Analýza klíčových slov na webu**

Funkce slouží k analýze klíčových slov ve struktuře a obsahu stránky. Pro všechna slova s výjimkou stop slov vypočte jejich výskyt v různých prvcích webu (titulek, H1-H3, popis, text, odkazy) a jejich hustotu.

## **Pozice ve vyhledávačích (na různá klíčová slova)**

Slouží k nalezení pozic ve vyhledávačích Seznam, Google CZ, Google EN a Google SK na zadaná klíčová slova a URL adresy. Hledaných URL adres lze zadat více najednou, a tudíž může sloužit i pro srovnání s konkurencí.

## **Nalezení zpětných odkazů**

V nabídce služeb je i vyhledání odkazujících domén, které ale nevypisuje žádné výsledky. Je to způsobeno tím, že k hledání využívají server search.yahoo.com, který již delší dobu neposkytuje informace o zpětných odkazech.

## **Hlídač zpětných odkazů**

Je to nástroj sloužící k ověření výskytu zpětného odkazu na odkazující stránce. Vhodný například pro hlídání placených odkazů nebo při výměně odkazů mezi partnerskými weby.

## **Hledanost frází**

Ukazuje statistiku hledanosti frází ve vyhledávači Seznam. Poskytuje průměrný počet hledání za den a týden, který je rozdělen ještě na rozšířenou a přesnou shodu.

## **Registrace do katalogů, tvorba microsites**

SeoNástroje nabízí hromadné vložení zpětných odkazů do katalogů. Tato služba je zpoplatněna částkou 240 Kč/100 katalogů.

## **SEO monitor**

Užitečná funkce, která automaticky měří a zjišťuje níže uvedené metriky a jejich report zasílá na e-mail uživatele v pravidelných intervalech:

- Hlídá pozice ve vyhledávačích (10 frází) – zobrazení rozdílů
- Počet indexovaných stránek
- Page Rank, S-rank

Nevýhodou je však nutnost přidání jejich zpětného odkazu na sledovaný web. Testováním jsem zjistil, že hodnoty se aktualizují jednou za dva dny, což při použití každodenního reportování může být matoucí.

## 3.3 SEO Power Suite

Jedná se o balíček 4 desktopových aplikací, kterými jsou Rank Checker na kontrolu pozic klíčových slov, WebSite Auditor na podrobnou analýzu webové stránky a SEO SpyGlass na zjištění zpětných odkazů a LinkAssistant sloužící především k vyhledání vhodných partnerských webů pro budování zpětných odkazů. Jelikož se moje práce zabývá zpětnými odkazy, popíšu jen poslední dvě jmenované.

### 3.3.1 LinkAssistant

Nástroj se zaměřuje na vyhledávání a správu partnerských webů sloužících k budování zpětných odkazů.

#### **Nalezení partnerských webů:**

Funkce nabízí vyhledání vhodných partnerských webů podle:

- Klíčových slov
- Webů odkazujících na konkurenci
- Webů odkazujících na můj web
- Sběrem všech URL na určité stránce
- Stránek s formulářem pro vložení odkazu (link submission)

Je možné nastavit, které metriky stránek se mají zjišťovat (Page Rank, Alexa Rank a Google Popularity). Aplikace hledá i kontaktní e-maily správců stránek, které později slouží ke komunikaci. Po nalezení seznamu odkazů se provádí ověření jejich existence. Zároveň se zjišťuje text odkazu a URL cílové stránky. Lze nastavit hloubku prohledávání nalezených webů.

#### **Odesílání e-mailů**

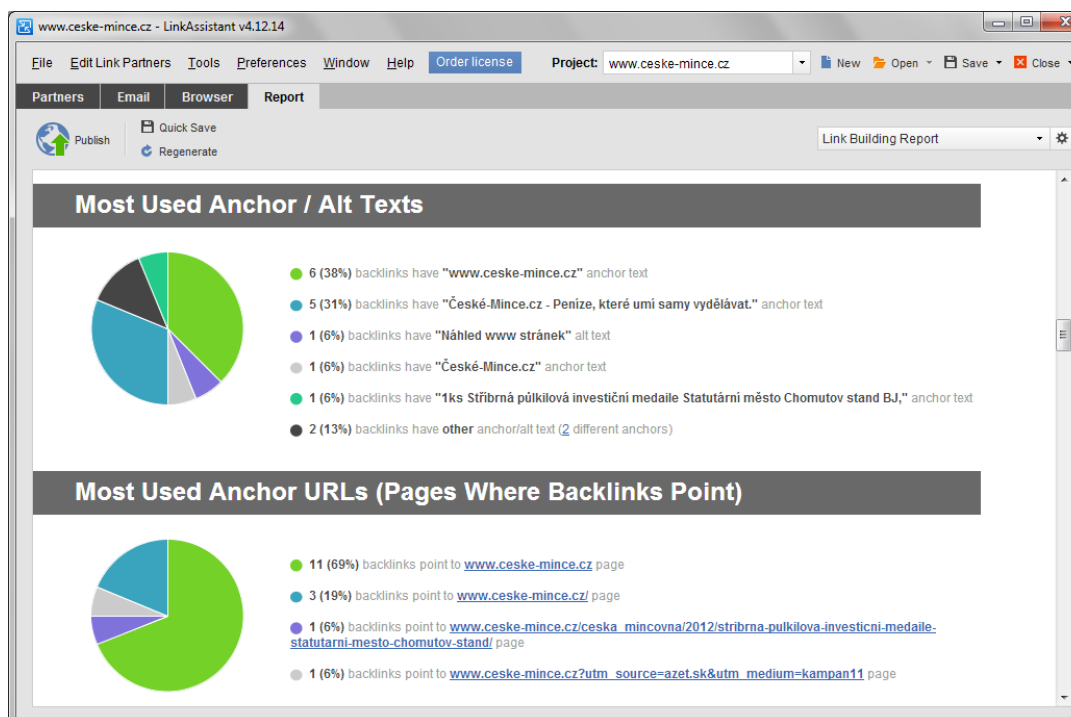
Aplikace umožňuje přímou komunikaci s partnery pomocí nalezených e-mailových kontaktů. K dispozici jsou vytvořené šablony e-mailů pro různé typy událostí.

#### **Vytvoření reportu**

Ze všech získaných údajů lze vygenerovat report současného stavu webu, který se skládá z:

- informací o doméně (Page Rank, Alexa Rank, Google Popularity)
- počtu zpětných odkazů a partnerů
- celkového link value
- Google Page Rank statistiky
- klíčových slov v textech
- textů odkazů
- cílových stránek
- výpisu odkazů se špatným textem nebo URL adresou

Reportovaná data jsou vizualizována pomocí grafů a tabulek, které jsou přehledně strukturovány do bloků (Obrázek 7).



Obrázek 7: Ukázka části reportu nástroje LinkAssistant

### 3.3.2 Seo SpyGlass

Má sloužit primárně ke zjištění zpětných odkazů od konkurentů, ale dá se využít i pro vlastní web. Pro konkrétní URL adresu se vyhledají zpětné odkazy a zobrazí se jejich seznam (v neplacené verzi 1000 odkazů). Poté lze seznam rozšířit o detailní informace odkazujících stránek.

#### Nalezení konkurenta a jeho zpětných odkazů

Aplikace nalezne největšího konkurenta podle zadaného klíčového slova nebo fráze a vyhledá jeho zpětné odkazy.

#### Zjištění zpětných odkazů pro konkrétní URL

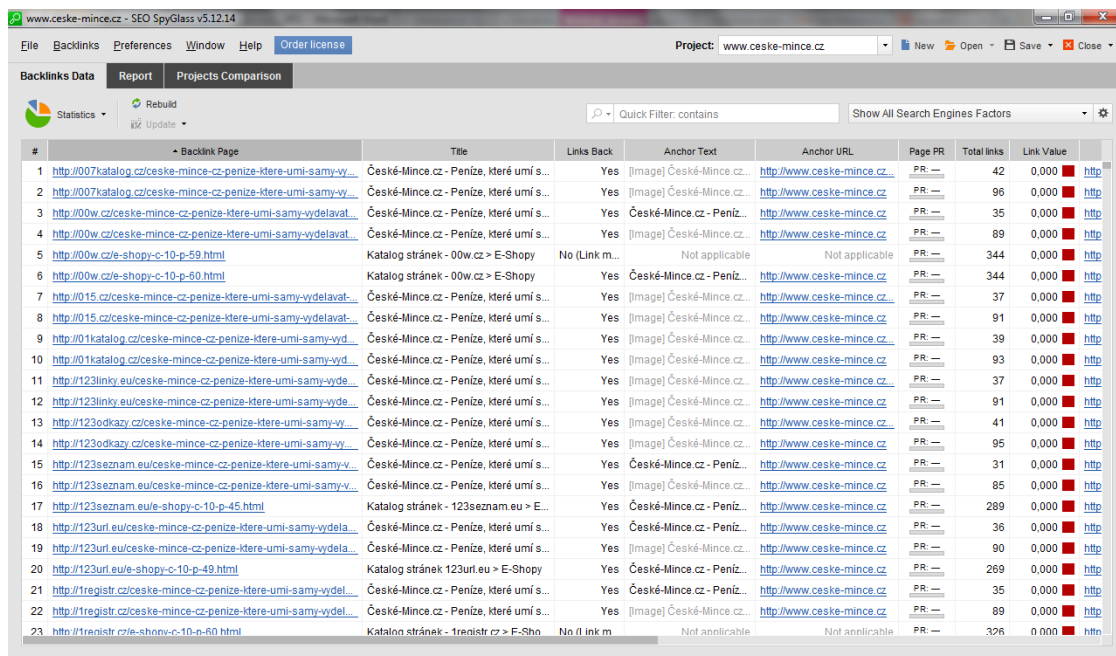
Vyhledá zpětné odkazy pro konkrétní webovou adresu.

## Analýza odkazů

Poté platí pro obě funkce stejný scénář. Po procesu vyhledání je k dispozici seznam URL adres odkazujících stránek, pro něž je možné zjistit doplňující informace, které jsou nezbytné k analýze odkazového profilu. Jsou to:

- Titulek stránky
- Text odkazu
- Cílová URL
- Page Rank
- Externí odkazy
- Celkem odkazy
- Link Value
- Doména
- IP adresa domény
- Page Rank domény
- Výskyt v DMOZ
- Alexa Rank
- Stáří domény

Proces stažení všech informací o zhruba 1500 odkazujících stránkách trval 27 minut a výsledek vypadá následovně (Obrázek 8).



The screenshot shows the 'Backlinks Data' report in the Seo SpyGlass application. The interface includes a menu bar (File, Backlinks, Preferences, Window, Help), a toolbar with 'Rebuild' and 'Update' buttons, and a search filter. The main table displays the following data:

#	Backlink Page	Title	Links Back	Anchor Text	Anchor URL	Page PR	Total links	Link Value
1	<a href="http://007katalog.cz/ceske-mince-cz-penize-ktere-umi-samy-vydelavat...">http://007katalog.cz/ceske-mince-cz-penize-ktere-umi-samy-vydelavat...</a>	České-Mince.cz - Peníze, které umí s...	Yes	[Image] České-Mince.cz...	<a href="http://www.ceske-mince.cz">http://www.ceske-mince.cz</a>	PR: —	42	0,000
2	<a href="http://007katalog.cz/ceske-mince-cz-penize-ktere-umi-samy-vydelavat...">http://007katalog.cz/ceske-mince-cz-penize-ktere-umi-samy-vydelavat...</a>	České-Mince.cz - Peníze, které umí s...	Yes	[Image] České-Mince.cz...	<a href="http://www.ceske-mince.cz">http://www.ceske-mince.cz</a>	PR: —	96	0,000
3	<a href="http://00w.cz/ceske-mince-cz-penize-ktere-umi-samy-vydelavat...">http://00w.cz/ceske-mince-cz-penize-ktere-umi-samy-vydelavat...</a>	České-Mince.cz - Peníze, které umí s...	Yes	České-Mince.cz - Peníz...	<a href="http://www.ceske-mince.cz">http://www.ceske-mince.cz</a>	PR: —	35	0,000
4	<a href="http://00w.cz/ceske-mince-cz-penize-ktere-umi-samy-vydelavat...">http://00w.cz/ceske-mince-cz-penize-ktere-umi-samy-vydelavat...</a>	České-Mince.cz - Peníze, které umí s...	Yes	[Image] České-Mince.cz...	<a href="http://www.ceske-mince.cz">http://www.ceske-mince.cz</a>	PR: —	89	0,000
5	<a href="http://00w.cz/e-shop-c-10-p-59.html">http://00w.cz/e-shop-c-10-p-59.html</a>	Katalog stránek - 00w.cz > E-Shopy	No (Link m...	Not applicable	Not applicable	PR: —	344	0,000
6	<a href="http://00w.cz/e-shop-c-10-p-60.html">http://00w.cz/e-shop-c-10-p-60.html</a>	Katalog stránek - 00w.cz > E-Shopy	Yes	České-Mince.cz - Peníz...	<a href="http://www.ceske-mince.cz">http://www.ceske-mince.cz</a>	PR: —	344	0,000
7	<a href="http://015.cz/ceske-mince-cz-penize-ktere-umi-samy-vydelavat...">http://015.cz/ceske-mince-cz-penize-ktere-umi-samy-vydelavat...</a>	České-Mince.cz - Peníze, které umí s...	Yes	[Image] České-Mince.cz...	<a href="http://www.ceske-mince.cz">http://www.ceske-mince.cz</a>	PR: —	37	0,000
8	<a href="http://015.cz/ceske-mince-cz-penize-ktere-umi-samy-vydelavat...">http://015.cz/ceske-mince-cz-penize-ktere-umi-samy-vydelavat...</a>	České-Mince.cz - Peníze, které umí s...	Yes	[Image] České-Mince.cz...	<a href="http://www.ceske-mince.cz">http://www.ceske-mince.cz</a>	PR: —	91	0,000
9	<a href="http://01katalog.cz/ceske-mince-cz-penize-ktere-umi-samy-vyd...">http://01katalog.cz/ceske-mince-cz-penize-ktere-umi-samy-vyd...</a>	České-Mince.cz - Peníze, které umí s...	Yes	[Image] České-Mince.cz...	<a href="http://www.ceske-mince.cz">http://www.ceske-mince.cz</a>	PR: —	39	0,000
10	<a href="http://01katalog.cz/ceske-mince-cz-penize-ktere-umi-samy-vyd...">http://01katalog.cz/ceske-mince-cz-penize-ktere-umi-samy-vyd...</a>	České-Mince.cz - Peníze, které umí s...	Yes	[Image] České-Mince.cz...	<a href="http://www.ceske-mince.cz">http://www.ceske-mince.cz</a>	PR: —	93	0,000
11	<a href="http://123linky.eu/ceske-mince-cz-penize-ktere-umi-samy-vyde...">http://123linky.eu/ceske-mince-cz-penize-ktere-umi-samy-vyde...</a>	České-Mince.cz - Peníze, které umí s...	Yes	[Image] České-Mince.cz...	<a href="http://www.ceske-mince.cz">http://www.ceske-mince.cz</a>	PR: —	37	0,000
12	<a href="http://123linky.eu/ceske-mince-cz-penize-ktere-umi-samy-vyde...">http://123linky.eu/ceske-mince-cz-penize-ktere-umi-samy-vyde...</a>	České-Mince.cz - Peníze, které umí s...	Yes	[Image] České-Mince.cz...	<a href="http://www.ceske-mince.cz">http://www.ceske-mince.cz</a>	PR: —	91	0,000
13	<a href="http://123odkazy.cz/ceske-mince-cz-penize-ktere-umi-samy-vy...">http://123odkazy.cz/ceske-mince-cz-penize-ktere-umi-samy-vy...</a>	České-Mince.cz - Peníze, které umí s...	Yes	[Image] České-Mince.cz...	<a href="http://www.ceske-mince.cz">http://www.ceske-mince.cz</a>	PR: —	41	0,000
14	<a href="http://123seznam.eu/ceske-mince-cz-penize-ktere-umi-samy-v...">http://123seznam.eu/ceske-mince-cz-penize-ktere-umi-samy-v...</a>	České-Mince.cz - Peníze, které umí s...	Yes	[Image] České-Mince.cz...	<a href="http://www.ceske-mince.cz">http://www.ceske-mince.cz</a>	PR: —	95	0,000
15	<a href="http://123seznam.eu/ceske-mince-cz-penize-ktere-umi-samy-v...">http://123seznam.eu/ceske-mince-cz-penize-ktere-umi-samy-v...</a>	České-Mince.cz - Peníze, které umí s...	Yes	České-Mince.cz - Peníz...	<a href="http://www.ceske-mince.cz">http://www.ceske-mince.cz</a>	PR: —	31	0,000
16	<a href="http://123seznam.eu/ceske-mince-cz-penize-ktere-umi-samy-v...">http://123seznam.eu/ceske-mince-cz-penize-ktere-umi-samy-v...</a>	České-Mince.cz - Peníze, které umí s...	Yes	[Image] České-Mince.cz...	<a href="http://www.ceske-mince.cz">http://www.ceske-mince.cz</a>	PR: —	85	0,000
17	<a href="http://123seznam.eu/e-shop-c-10-p-45.html">http://123seznam.eu/e-shop-c-10-p-45.html</a>	Katalog stránek - 123seznam.eu > E...	Yes	České-Mince.cz - Peníz...	<a href="http://www.ceske-mince.cz">http://www.ceske-mince.cz</a>	PR: —	289	0,000
18	<a href="http://123url.eu/ceske-mince-cz-penize-ktere-umi-samy-vyde...">http://123url.eu/ceske-mince-cz-penize-ktere-umi-samy-vyde...</a>	České-Mince.cz - Peníze, které umí s...	Yes	České-Mince.cz - Peníz...	<a href="http://www.ceske-mince.cz">http://www.ceske-mince.cz</a>	PR: —	36	0,000
19	<a href="http://123url.eu/ceske-mince-cz-penize-ktere-umi-samy-vyde...">http://123url.eu/ceske-mince-cz-penize-ktere-umi-samy-vyde...</a>	České-Mince.cz - Peníze, které umí s...	Yes	[Image] České-Mince.cz...	<a href="http://www.ceske-mince.cz">http://www.ceske-mince.cz</a>	PR: —	90	0,000
20	<a href="http://123url.eu/e-shop-c-10-p-49.html">http://123url.eu/e-shop-c-10-p-49.html</a>	Katalog stránek 123url.eu > E-Shopy	Yes	České-Mince.cz - Peníz...	<a href="http://www.ceske-mince.cz">http://www.ceske-mince.cz</a>	PR: —	269	0,000
21	<a href="http://1registr.cz/ceske-mince-cz-penize-ktere-umi-samy-vyde...">http://1registr.cz/ceske-mince-cz-penize-ktere-umi-samy-vyde...</a>	České-Mince.cz - Peníze, které umí s...	Yes	České-Mince.cz - Peníz...	<a href="http://www.ceske-mince.cz">http://www.ceske-mince.cz</a>	PR: —	35	0,000
22	<a href="http://1registr.cz/ceske-mince-cz-penize-ktere-umi-samy-vyde...">http://1registr.cz/ceske-mince-cz-penize-ktere-umi-samy-vyde...</a>	České-Mince.cz - Peníze, které umí s...	Yes	[Image] České-Mince.cz...	<a href="http://www.ceske-mince.cz">http://www.ceske-mince.cz</a>	PR: —	89	0,000
23	<a href="http://1registr.cz/e-shop-c-10-p-60.html">http://1registr.cz/e-shop-c-10-p-60.html</a>	Katalog stránek - 1registr.cz > F-Sho...	No (Link m...	Not applicable	Not applicable	PR: —	326	0,000

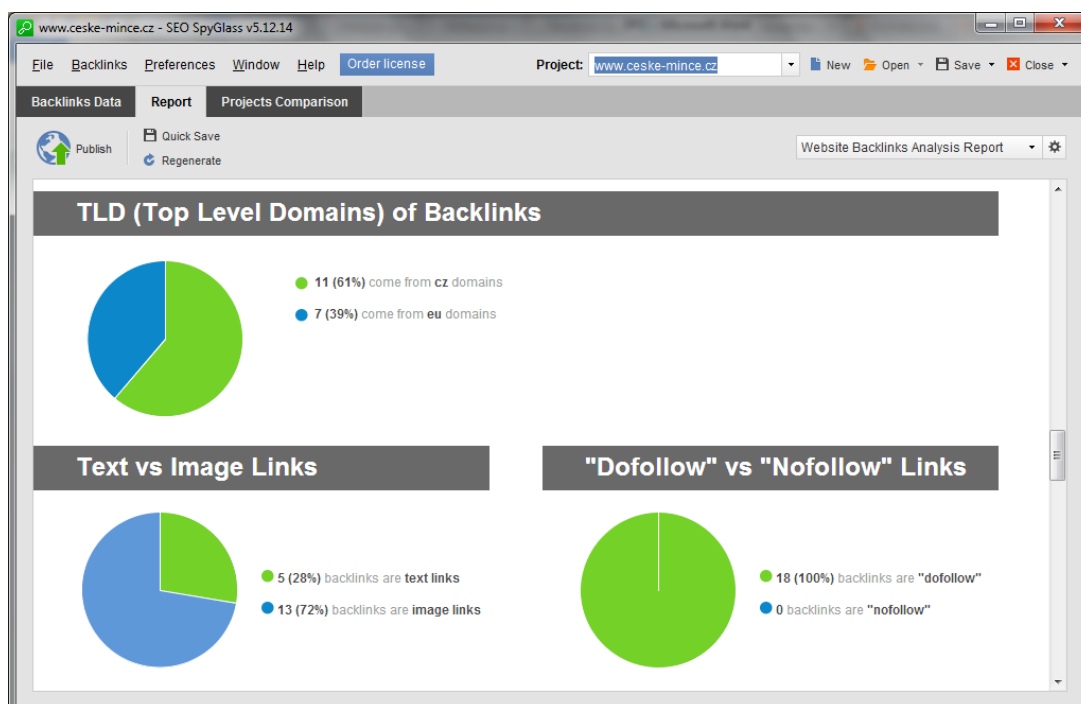
Obrázek 8: Ukázka nástroje Seo SpyGlass

Po získání dat o odkazujících stránkách si uživatel může zobrazit analýzu odkazového profilu, která je složena z následujících veličin:

- Seznam a četnost domén nejvyšší úrovně (TLD)
- Seznam a četnost domén
- Četnost Page Rank hodnot
- Seznam klíčových slov s hustotou a četností v titulku a textu odkazu
- Seznam a četnost textů odkazů
- Seznam a četnost cílových URL
- Seznam a četnost IP adres
- Seznam a četnost odkazujících zemí
- Výskyt stránek v DMOZ

## Reportů

Ze seznamu odkazů a jejich metrik lze sestavit souhrnné reporty aktuálního stavu odkazového profilu (Obrázek 9). Stejně jako u LinkAssistant jsou velmi přehledné, čitelné a nabízí dostatečné množství informací.



Obrázek 9: Ukázka reportu SEO SpyGlass

## 3.4 WebCEO

Jedná se o velice komplexní zahraniční nástroj pro SEO analýzu. Existuje jak online, tak i v desktopové verzi, která je zdarma k vyzkoušení na jeden měsíc. Nabízí funkce na zjištění pozic klíčových slov pro různé vyhledávací enginy. Sleduje aktivitu na sociálních sítích jako je počet sdílení a komentářů na serveru Facebook, počet +1 na Google Plus a počet příspěvků na Twitter. Dále nabízí SEO audit on-page faktorů webu, špehování konkurence z pohledu pozic na klíčová slova, počtu odkazů atd.

Uživatel si může na hlavní souhrn umísťovat různé panely s náhledy analýz, které po prokliku nabízí detailní informace. Aplikace je rozdělena do různých cenových kategorií. Bezplatné a levnější verze jsou omezeny počtem projektů, klíčových slov, konkurenčních webů, počtem zobrazených odkazů, reporty aj.

### 3.4.1 Funkce

#### **Klíčová slova**

Funkce nabízí měření pozic klíčových slov v několika vyhledávacích. U verze zdarma je možné analyzovat pouze 5 klíčových slov pro 5 různých vyhledávacích enginů. Lze sledovat i průměrné pozice vybraných konkurenčních webů (max. 20 v bezplatné verzi). Zjišťuje se:

- Hledanost slova/fráze za měsíc
- Pozice (aktuální + porovnání s minulou)
- Nejlepší webová stránka na dané slovo/frázi
- Pozice konkurentů

#### **Audit webu**

Poskytuje kompletní audit webu zahrnující:

- Kontrolu odkazů (interní, externí)
- Chybějící klíčová slova na stránkách
- SEO problémy
  - Chybějící TITLE, DESCRIPTION, H1
  - Duplicitní TITLE, DESCRIPTION
  - Délka TITLE, DESCRIPTION
  - Špatné přesměrování
  - Hustota klíčových slov v H1 a odchozích odkazech
  - Struktura URL adres
  - Přítomnost XML Sitemap a Robots.txt

## Sociální sítě

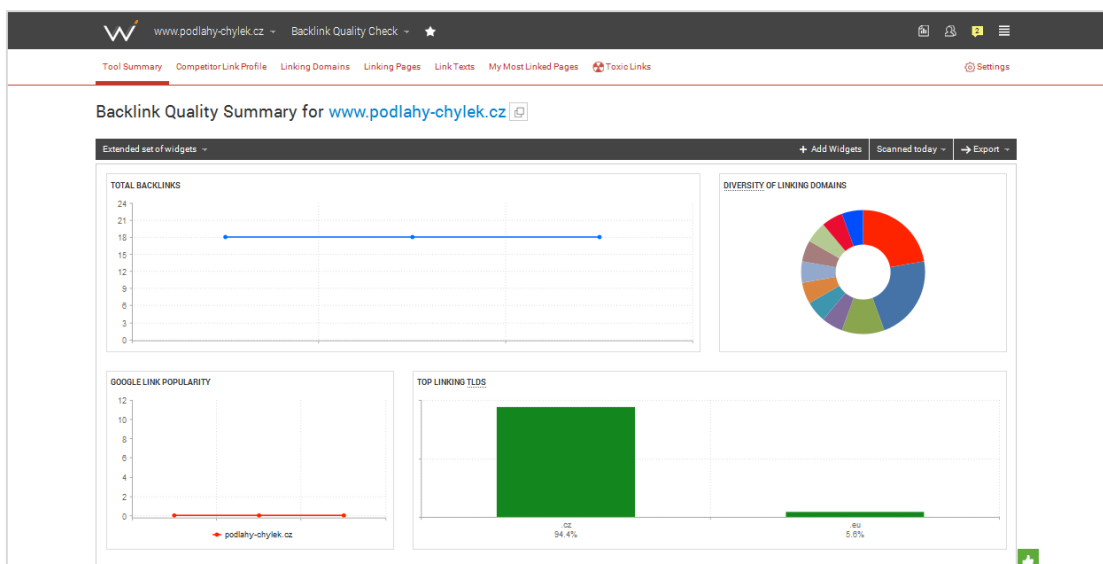
Zjišťuje počet výskytů URL adresy na sociálních sítích:

- Facebook – sdílení, likes, komentáře
- Google +1
- Twitter

## Zpětné odkazy

Nástroj nabízí komplexní analýzu odkazového profilu s detailními informacemi o všech metrikách. Jednotlivé analýzy jsou mezi sebou aktivně propojeny, což poskytuje získání informací z různých pohledů.

Na titulní straně je k dispozici shrnutí analýz odkazového profilu formou grafů a tabulek (Obrázek 10). Jejich zobrazení a pozice je volitelné a lze měnit pomocí „drag and drop“.



Obrázek 10: Ukázka nástroje Web CEO – shrnutí analýzy odkazů

Shrnutí nabízí například celkový počet odkazujících stránek, domén a IP, graf Google Popularity, graf s počtem výskytů jednotlivých domén, TLD, textů odkazů aj.

Další záložky odkazového profilu nabízí informace o:

- Odkazech konkurentů
- Odkazujících doménách
- Odkazujících stránkách
  - URL odkazující stránky
  - Velikost stránky
  - Počet vnitřních odkazů
  - Počet vnějších odkazů
  - Alexa Traffic Rank
  - IP adresa
  - Anchor text
  - Cílová stránka
  - Google PR
  - Google PR domovské stránky
  - Status (OK, nofollow, toxic)
- Textech odkazů
- Cílových stránkách odkazů

### 3.4.2 Tarify a omezení

Nabízí velké množství různých tarifů, jejichž cena se odvíjí podle maximálního počtu zkoumaných veličin (Tabulka 3).

	FREE	SILVER	GOLD	PLATINUM	PRO
<b>Počet projektů</b>	2	5	10	25	2\$ / projekt
<b>Klíčových slov</b>	5	100	300	750	Neomez.
<b>SERP stran</b>	3	3	3	5	10
<b>Zpětné odkazy</b>	40	5000	30000	125500	Neomez.
<b>Konk. zpětné odkazy</b>	0	7500	25000	62500	Neomez.
<b>Reporty</b>	ne	týdně	týdně	týdně	denně
<b>Cena/měsíc (Kč)</b>	0	1380	2980	6980	1980+

Tabulka 3: Tarify a omezení u Web CEO



## 3.5 Shrnutí

Collabim je dobrý nástroj pro měření pozic ve vyhledávačích na různá klíčová slova. Nicméně práce s ním je místy dost nepřehledná a uživatelsky nepřívětivá, zejména u přidávání klíčových slov a navigace v systému. Velkým nedostatkem je absence nástroje na vyhledávání zpětných odkazů.

SeoNástroje neposkytují tak kvalitní služby, jako ostatní zkoumané nástroje, ale nabízí pár zajímavých funkcí jako například SEO monitor, který zasílá pravidelné reporty aktuálního stavu nebo ruční registrace do katalogů. Je vhodný spíše pro analýzu on-page faktorů a měření pozic klíčových slov.

Web CEO je velmi komplexní nástroj nabízející možnost celkové analýzy webu. V sekci zpětných odkazů využívá databázi od ahrefs.com, která má vlastní vyhledávací roboty prohledávající internet. Dokáže tak poskytnout výsledky téměř ihned. Na druhou stranu mu trvá déle najít nově vzniklé odkazy, neboť některé stránky jsou pravidelně procházeny roboty méně než jiné.

SEO SpyGlass podávalo dobré výsledky, co se týče počtu odkazů a informací o nich, ale nevýhodou bylo, že se mezi výsledky objevovalo i množství již neaktivních odkazujících stránek. Celkově tento nástroj ztrácel na neaktuálnosti odkazového profilu. Odhalit nově přidané odkazy trvá delší dobu. Také chyběla informace o typu odkazu (follow, nofollow, 301 atd.). U některých menších webů s nízkým počtem odkazů nenašel žádné výsledky.

LinkAssistant je nástroj použitelný pro nalezení nových link partnerů. Nevýhodou nástroje je absence češtiny, která vadí zejména při zadávání klíčových slov. Také chybí možnost uložení nalezených dat u bezplatné verze. Při použití funkce pro nalezení odkazujících webů neposkytuje příliš dobré výsledky.

## 4 Nástroj pro analýzu a budování odkazového profilu

Studium různých SEO technik a triků vhodných pro budování zpětných odkazů mě přivedlo k myšlence vytvořit univerzální nástroj, který bude sloužit jak k zjištění stávajícího stavu odkazového profilu, tak i k evidenci nově vytvořených odkazů. Zároveň bude analyzovat nejrůznější metriky a sloužit k realizaci nejlepších SEO praktik. Například při psaní článků je důležitá jejich unikátnost napříč celým internetem a není možné si různé varianty pamatovat. Proto je užitečné si texty článků i anchor textů evidovat. Potřebná je i kontrola zpětných odkazů, zejména těch placených. Nedávné úpravy hodnotících algoritmů si žádají změny v dosavadních praktikách budování zpětných odkazů. Je potřeba zohledňovat více kritérií, což nelze provádět bez komplexního nástroje.

Vytvořený nástroj je součástí společného projektu, který vznikl v rámci diplomových a bakalářských prací. Jedná se o komplexní analytický SEO nástroj složený z několika modulů, které používají společné jádro aplikace, ale jsou na sobě nezávislé. Mezi funkcemi společného projektu jsou SEO analyzátor (on-page faktory), agregátor cen produktů, měřič pozic ve vyhledávacích, analyzátor e-shopů a již zmíněný analyzátor zpětných odkazů.

Při tvorbě aplikace jsem vycházel z analýzy konkurenčních nástrojů a jejich funkcí. Využil jsem také poznatky získané při studiu nejlepších SEO postupů a některé z nich jsem následně promítl do funkcionalit svého systému. Snažil jsem se také zohlednit nejdůležitější požadavky ze strany vyhledávacích enginů a jejich hodnotících algoritmů.

### 4.1 Funkční analýza

#### 4.1.1 Funkce systému

- Stahování dat o odkazech
- Analýza odkazového profilu webu
- Tvorba pravidelných reportů
- Vyhledání tematických webů
- Budování zpětných odkazů

#### 4.1.2 Vstupní data

- Uživatelské údaje
- Údaje o projektu
- Seznam URL adres projektových webů (včetně konkurentů)
- Seznam klíčových slov k projektu
- Firemní údaje potřebné k registraci do katalogů
- Data potřebná k registraci do katalogů – titulky, popisy, články

### 4.1.3 Výstupy ze systému

- Seznam odkazujících stránek
- Seznam odkazujících domén
- Texty zpětných odkazů
- Seznam cílových stránek
- Srovnání s konkurenty
- Seznam registrovaných katalogů
- Seznam katalogů vhodných k registraci
- Seznam katalogů připravených k registraci
- Seznam plánů na budování zpětných odkazů
- Graf počtu odkazujících domén, stránek, IP adres a Google odkazů
- Graf domén nejvyšší úrovně
- Graf hodnocení webu (Page Rank, S-rank, Alexa rank)
- Tabulka odkazů z důležitých domén
- Tabulka s počtem výskytů v sociálních sítích

### 4.1.4 Uživatelé systému

Systém slouží majitelům a tvůrcům webových stránek, kteří si mohou vytvářet uživatelské účty, projekty a prohlížet analýzy.

- Administrátor
  - všechny uživatelské funkce
  - neomezený přístup ke všem datům
  - manuální spouštění stahovacích procesů
- Uživatel
  - vytvoření a editace uživatelského účtu
  - vytvoření a editace projektu
  - vkládání, editace a mazání klíčových slov, produktů a konkurentů
  - zobrazení analýz a dat u svých projektů

## 4.2 Databázový model

Datová vrstva aplikace je postavena na databázovém systému MySQL. Jako formát úložiště dat je použit InnoDB, protože umožňuje práci s cizími klíči a podporuje transakce. Při tvorbě datového modelu se dbalo nejen na omezení redundance dat, ale i na rychlost dotazování. Z toho důvodu byly do některých tabulek přidány pomocné cizí klíče, které zredukovaly použití operací typu JOIN, a tak zvýšily rychlost provádění dotazů. U tabulek, kde se často používají operace na vyhledávání a třídění jsou dané sloupce vhodně indexovány. Jelikož je moje aplikace jedním z mnoha modulů společného projektu, využívá i společné tabulky. V této práci však popíšu jen tabulky vztahující se k mému modulu, tedy ty, které jsem sám vytvářel, a jednu společnou, která je zásadní pro chod aplikace. Návrh databáze je zobrazen ER-diagramem viz Obrázek 11.

## 4.2.1 Datový slovník

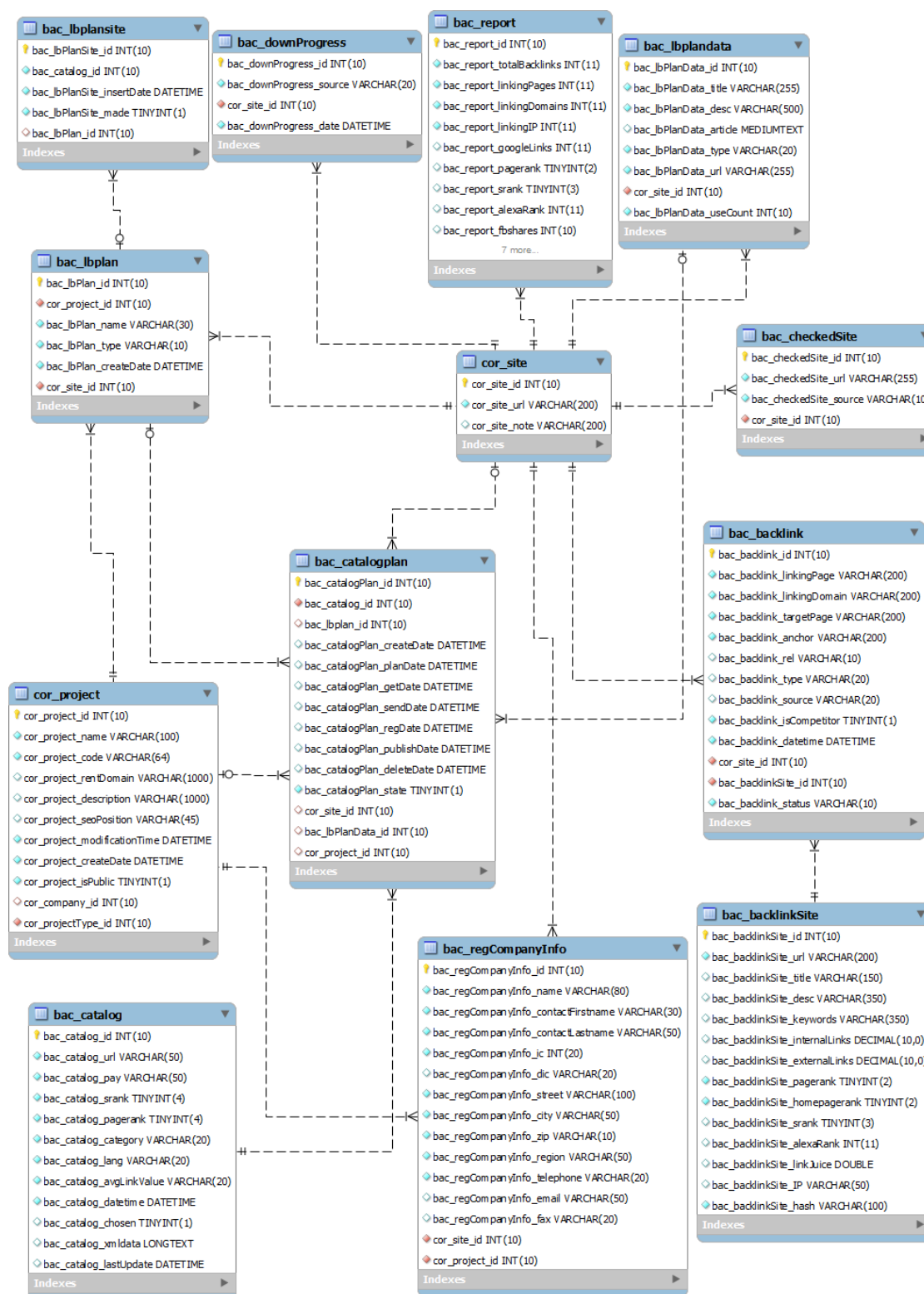
Pro výpis databázových entit jsem použil datový slovník, který zahrnuje název atributu, datový typ, velikost, informaci zda může být NULL a zda se jedná o primární nebo cizí klíč. Tento seznam je součástí přílohy (Datový slovník).

Ukázka datového slovníku tabulky BAC\_BACKLINKSITE, která shromažďuje data o odkazujících stránkách (Tabulka 4):

BAC_BACKLINKSITE			
Atribut	Datový typ	NULL	Klíč
bac_backlinkSite_id	int(10) <i>Auto Increment</i>	Ne	PK
bac_backlinkSite_url	varchar(200)	Ne	Ne
bac_backlinkSite_title	varchar(150)	Ano	Ne
bac_backlinkSite_desc	varchar(350)	Ano	Ne
bac_backlinkSite_keywords	varchar(350)	Ano	Ne
bac_backlinkSite_internalLinks	decimal(10,0)	Ne	Ne
bac_backlinkSite_externalLinks	decimal(10,0)	Ne	Ne
bac_backlinkSite_pagerank	tinyint(2)	Ne	Ne
bac_backlinkSite_homepagerank	tinyint(2)	Ne	Ne
bac_backlinkSite_srank	tinyint(3)	Ne	Ne
bac_backlinkSite_alexRank	int(11)	Ne	Ne
bac_backlinkSite_linkJuice	double	Ne	Ne
bac_backlinkSite_IP	varchar(50)	Ne	Ne
bac_backlinkSite_hash	varchar(100)	Ne	Ne

Tabulka 4: Datový slovník tabulky bac\_backlinkSite

### 4.2.2 ER-Diagram



Obrázek 11: ER-diagram databáze modulu

## 4.3 Popis implementace

Velké konkurenční projekty používají algoritmy zvané „crawlers“, které prohledávají miliony webových stránek, ze kterých extrahují veškeré jejich odkazy. Tímto způsobem jsou schopny mapovat prostředí internetu a vazby mezi stránkami, tedy i zpětné odkazy. Avšak toto řešení vyžaduje velké množství prostředků a výpočetního výkonu.

### 4.3.1 Proces hledání zpětných odkazů

Největším úskalím a zároveň výzvou této práce bylo vymyslet způsob, jakým zjistit aktuální a pokud možno kompletní odkazový profil určité webové stránky. Jelikož jsem neměl potřebné množství výpočetního výkonu, abych mohl procházet stovky milionů stránek a stahoval jejich odkazy, musel jsem nalézt jinou metodu. V této kapitole se budu věnovat popisu technik a zdrojů, které jsem využil pro agregaci zpětných odkazů.

V začátcích své práce jsem se snažil hledat zpětné odkazy pomocí dolování dat z existujících webových služeb jako například opensiteexplorer.com, ahrefs.com, webceo.com a jiné. Všechna tato řešení si však pečlivě střeží svá data a neplatícím uživatelům poskytují jen zlomek skutečných odkazů. Navíc ani parsování těchto stránek není jednoduché, neboť používají řadu obranných mechanismů. Patří mezi ně omezení počtu požadavků z jedné IP adresy, generování speciálních HASH kódů v parametru POST nebo vyžadování registrace a přihlášení. Po pečlivém zkoumání podobných řešení jsem dospěl k názoru, že touto cestou se nelze dále ubírat a to ze dvou důvodů. Prvním byl fakt, že v neplacené verzi neposkytují dostačující služby a nelze je z nich snadno dolovat. Druhým, a podle mého názoru důležitějším důvodem, bylo omezení závislosti na cizích řešeních, což by z dlouhodobého hlediska mohlo přinášet problémy. Usoudil jsem, že nebude možné spoléhat se pouze na jeden zdroj dat, a tak bylo nutné vymyslet a otestovat alternativní způsoby hledání odkazujících stránek a efektivně je kombinovat do jednoho celku.

### Google vyhledávač

Začal jsem s vyhledáváním zpětných odkazů podle URL adresy webu ve vyhledávači Google. Prvotní tvar dotazu vypadal následovně –  $q=www.domena.cz$ . Problém nastal v případě, když se v názvu domény objevovala skutečná slova. Vyhledávací engine tato slova izoloval a zobrazil k nim výsledky, které však nebyly relevantní k hledané adrese. Pro zpřesnění výsledků jsem adresu uvedl jako přesný výraz s použitím dvojitého uvozovky:  $q="www.domena.cz"$ . Seznam výsledných stránek se značně omezil a přiblížil se očekávanému cíli. Protože se ve výsledcích objevovaly i stránky z analyzovaného webu, bylo zapotřebí použít speciální operátor  $-site:www.domena.cz$ , který z výsledků vyloučí zadanou doménu. Může nastat situace, že se webová adresa na internetu vyskytuje i bez subdomény (např.  $www$ ), a proto je nutné s tímto případem počítat. Výsledný tvar vyhledávacího dotazu tedy vypadá následovně:

$$q="domena.cz" -site:domena.cz$$

Takto byly nalezeny první odkazující webové stránky na zadanou doménu. Jejich počet však zdaleka neodpovídal reálnému počtu referujících webů a bylo potřeba jej rozšířit o další zdroje.

Internetové stránky se často v textech odkazují pod svým názvem nebo značkou tzv. brandovou frází. Tímto slovním spojením uvedeným ve vyhledávacím dotazu bylo možné nalézt další řadu odkazujících stránek.

$$q = \text{"brand" -site:domena.cz}$$

Další variantou jak najít zpětné odkazy je přidání klíčového slova do vyhledávací fráze. Výsledky v tomto případě nejsou tak přesné, jako u předchozích dvou způsobů, ale kombinací více klíčových slov je po menších počtech možné vyhledat pár odkazujících stránek. Seznam výsledků se dá opět zúžit přidáním domény v uvozovkách.

$$q = \text{klíčové slovo "domena.cz" -site:domena.cz}$$

### **Google operátor link:**

Je to speciální operátor používaný při vyhledávání, který má najít stránky odkazující na daný web uvedený za dvojtečkou. V minulosti se jednalo o velmi používaný způsob hledání zpětných odkazů, protože Google přes něj poskytoval celkem kvalitní výsledky. V současnosti, kdy Google zbrojí proti nekvalitním odkazům, razantně změnil výsledky této služby a zobrazuje pouze zlomek odkazujících webů, což službu zařadilo mezi nerelevantní zdroje. Využívá se jako měřítko tzv. Google Link Popularity, což je počet výsledků zjištěných právě tímto operátorem. Odkazy nalezené tímto způsobem se dají považovat za kvalitnější.

$$q = \text{link:domena.cz}$$

### **Problém s penalizací**

Při testování a zjišťování nejlepších postupů docházelo k častému dotazování ze stejné IP adresy, což u vyhledávače vzbudilo pozornost. Jelikož se vyhledávání neprovádělo ručně, ale pomocí skriptu, došlo ve velmi krátké době k jeho odhalení vedoucímu k neustálým blokacím a nutností ověření pomocí CAPTCHA kódu. Snažil jsem se tento problém vyřešit nastavením náhodných prodlev mezi dotazy a změnou typů prohlížečů, což sice fungovalo, ale jen krátce. Vyzkoušel jsem možnost různých PROXY serverů, ale nenašel jsem stabilní řešení vhodné pro můj projekt. Bez většího počtu měnících se IP adres nebylo možné klasické vyhledávání využít, a proto jsem se uchýlil k využití Google Custom Search API.

### **Google Custom Search API (CSE)**

Společnost Google nabízí službu Vlastní vyhledávání, kterou lze integrovat do webové stránky nebo ji využívat pomocí API rozhraní. V neplacené verzi, která je pro tento projekt využita, je k dispozici 100 dotazů za den. Jeden dotaz se rovná jedné stránce s výsledky vyhledávání. Na jednu vyhledávací frázi lze získat maximálně 100 výsledků, tedy 10 stránek.

### **Vyhledávání na specifických doménách:**

API nabízí kromě vyhledávání na celém internetu i možnost vyhledávat jen na zadaných doménách. Omezit výsledky jen na určitou doménu funguje i pomocí operátoru „site:“ v běžném dotazu. Toto řešení však ve srovnání s vyhledáváním omezeným pouze na jednu doménu poskytuje mnohem méně výsledků a v určitých případech dokonce žádné, i když se tam odkaz nachází. Bylo vytvořeno několik druhů speciálních vyhledávačů s omezením jen na domény, které mají na českém internetu stěžejní význam. Jednalo se o:

- \*.firmy.cz
- \*.zbozi.cz
- \*.aukro.cz
- \*.heureka.cz
- \*.centrum.cz
- \*.adresarfirem.cz

Výsledky hledání byly o poznání lepší a díky tomuto řešení byla aplikace schopna najít mnohem více odkazujících stránek. Stejné nastavení bylo využito i pro vyhledávání v seznamu katalogů k registraci.

### **Prohledávání katalogů**

Pro vyhledání odkazového profilu v katalogích byla implementována funkce na procházení každého katalogu zvlášť. Pro tento postup bylo nejprve použito CSE s aplikováním dotazu „doména.cz“ *site:katalog.cz*. Průběžné testování však odhalilo stejný problém, jaký je popsán o odstavec výše. Bylo tedy nutné vytvořit sadu upravených vyhledávačů tak, aby byla zajištěna rovnováha mezi efektivitou při hledání odkazujících katalogů a potřebným počtem vyhledávačů. Nejprve byla vyzkoušena možnost přidání všech katalogových domén do jednoho vyhledávacího enginu, což poskytovalo výsledky téměř shodné s operátorem „site:“. Počet katalogů tedy bylo zapotřebí snížit na optimální hodnotu, která by se ideálně měla blížit číslu jedna. Vytvoření tak velkého počtu vlastních vyhledávačů však nebylo reálné. Výsledky začaly být relevantní při hodnotě 10 katalogů v jednom vyhledávači. Pro zajištění možnosti prohledání všech katalogů najednou i při omezení 100 dotazů na den, bylo v jednom CSE účtu vytvořeno 6 vyhledávačů a každý z nich vyhledával na 12 katalogových doménách.

### **SEOmoz API**

Jako doplňková metoda bylo použito API od společnosti MOZ.com, které nabízí registrovaným uživatelům mnoho zajímavých služeb pro zjišťování informací o webových stránkách a jejich odkazech. Pro účely této práce byla využita funkce na vyhledávání zpětných odkazů. Vstupem byla URL adresa webu, přidělený API klíč a správné nastavení parametrů:



- **Scope** – udává, jaký typ odkazu se má vracet (page, domain, subdomain)
- **Sort** – řazení podle
  - autority stránky (*page\_authority*)
  - autority domény (*domain\_authority*)
  - podle počtu domén směřujících na doménu (*domains\_linking\_domain*)
  - podle počtu domén směřujících na stránku (*domains\_linking\_page*)
- **Filter** – filtrování podle typu odkazu (*internal, external, follow, nofollow, 301, 302*); filtry je možné kombinovat pomocí operátoru „+“
- **Limit and Offset** – omezení počtu výsledků (základní hodnota je 25 a nejvyšší 50)
- **TargetCols** – číslo specifikující, která data o cílových stránkách budou na výstupu
- **SourceCols** – číslo specifikující, která data o odkazujících stránkách budou na výstupu
- **LinkCols** – číslo specifikující, která data o odkazech budou na výstupu

[9]

Výstupem byly informace o odkazujících stránkách a odkazech ve formátu JSON.

```
"[{ "lrid":366183714564,
  "lsrc":59276924485,
  "ltgt":59407312237,
  "luuu":"podlahy-chylek.cz/",
  "uu":"www.dvereinterier.kvalitne.cz/" } ]
```

Struktura je převedena na pole, ze kterého se získá seznam URL adres odkazujících stránek, připravený k dalšímu procesu zpracování.

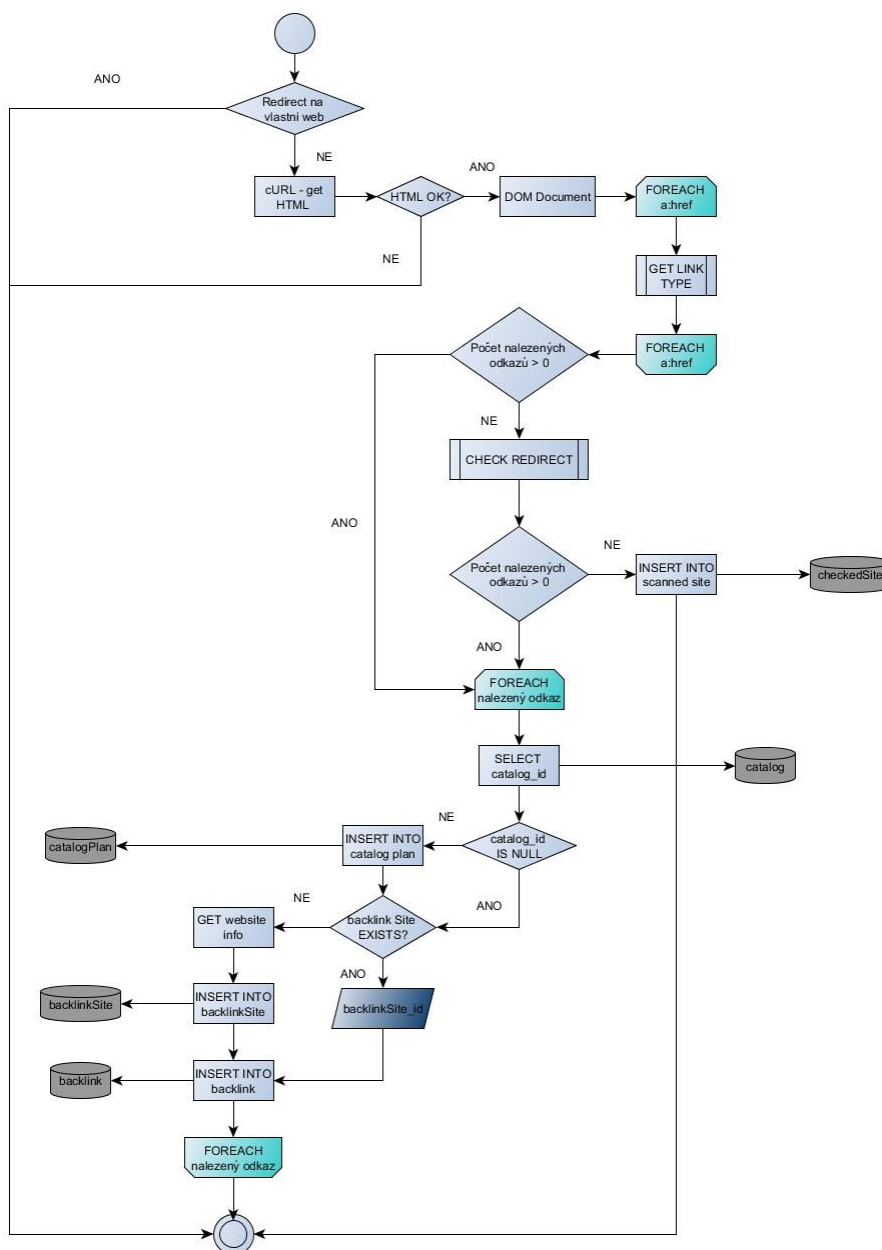
Bylo provedeno množství testů k odhalení optimálního nastavení vstupních parametrů. Pro získání co největšího počtu odkazů se vzájemně kombinují hodnoty vstupů scope, sort a filter. Speciálně u parametru filter se navíc tvoří dvojice pomocí znaménka „+“ (external+nofollow, external+301 atd.). API mohlo vracet pro různé kombinace podobné nebo stejné výsledky, a proto byla implementována kontrola duplicity, která zajišťuje, že jednou prohledaná stránka bude při dalším nálezu ignorována. Služba je omezena na 1 milion vrácených položek za měsíc, což pro účely projektu bohatě stačí.

## 4.3.2 Parsování odkazujících stránek

### Nalezení zpětného odkazu

Po vyhledání odkazující stránky jednou z výše uvedených metod se spouští proces nalezení zpětného odkazu v jejím kódu. Vstupy pro tento proces jsou URL adresa, doména hledaného webu a její ID. Nejprve se HTML stránka uloží jako objekt typu DomDocument, pomocí kterého lze přistupovat k jednotlivým elementům. Poté se pro každý odkaz (element „a“) zjistí jeho atribut „href“, jehož tvar se porovná s hledanou doménou. Tvary obou porovnávaných adres musí být reprezentovány doménami 1. a 2. úrovně. Jakmile je nalezena shoda, zjistí se text odkazu a jeho atribut „rel=“ a následně se vše uloží do pole nalezených odkazů. V opačném

případě je odkaz zařazen do pole interních nebo externích odkazů podle toho, jestli směřuje dovnitř nebo ven z webu. Po kontrole všech odkazů se projde pole nalezených. Nejprve se podle domény prohledá seznam katalogů. V případě nálezu se vytvoří nebo aktualizuje záznam v plánu pro budování odkazů. Následně dojde ke zjištění potřebných dat o odkazující stránce a jejímu uložení do databáze. Pokud již stránka existuje, vezme se pouze její ID a přiřadí se k odkazu při jeho uložení. Za předpokladu, že nebyl nalezen žádný zpětný odkaz, se odkazující stránka uloží do tabulky zkontrolovaných a při příštím nálezu bude ignorována. Algoritmus je názorně popsán vývojovým diagramem (Obrázek 12).



Obrázek 12: Vývojový diagram nalezení zpětného odkazu

## Informace o odkazující stránce

K provedení komplexní analýzy nalezeného zpětného odkazu je zapotřebí zjistit co nejvíce informací o odkazující stránce. Mezi hlavními sledovanými metrikami jsou ty, které určují hodnotu odkazů:

- počet interních a externích odkazů
- Google Page Rank
- Seznam S-rank

Z těchto údajů získáme za použití vzorce tzv. link juice, který udává kvalitu referenční stránky. K určení míry relevance webu se eviduje:

- Titulek stránky
- Popis stránky
- Klíčová slova

V případě, že některý z těchto atributů obsahuje klíčová slova vztahující se k odkazovanému webu, může mu to přinést lepší hodnocení a odkaz se tak stává kvalitnějším.

## Informace o odkazu

Na jedné odkazující stránce se může vyskytovat více odkazů směřujících na jednu doménu. Mohou se lišit svým typem nebo vzhledem. Existují v klasické textové formě, jako obrázek nebo jsou obklopeny jiným elementem například tlačítkem. Ve vyhledávacím algoritmu tento typ zjišťují a evidují pro pozdější analýzu. Například mohou určit, kolik procent odkazů pochází z obrázků. Nejdůležitějším atributem je samozřejmě text odkazu, který by měl být vhodně zvolený a obsahovat relevantní klíčové slovo. Evidovanými atributy jsou:

- Text odkazu
- Cílová stránka
- Atribut „rel“ (dofollow, nofollow, 301, 302)
- Typ katalogu (pokud existuje)

Pro rozpoznání, zda zpětný odkaz opravdu směřuje na hledaný web, je potřeba mít obě URL adresy ve stejném tvaru. Adresa zdrojového odkazu i cílové stránky se porovnává podle domény 2. a 3. řádu. Pro tyto účely jsem implementoval funkci, která jakoukoliv URL adresu vrátí ve tvaru domény a subdomény.

## Detekce přesměrování

Mnoho serverů, zejména katalogů, odkazuje na registrované weby pomocí přesměrování a zobrazená URL není ve tvaru, v jakém ji hledáme. Pro odhalení těchto zpětných odkazů je zapotřebí kontrolovat jednotlivé URL a zjišťovat jejich cílovou adresu po přesměrování. Využívám k tomu knihovnu cURL, pomocí které zjistím skutečnou cílovou adresu. Takto objevené odkazy jsou označeny typem přesměrování, aby mohly být zpracovány příslušnou

analýzou. Je zřejmé, že tímto postupem se zvyšuje výpočetní výkon celého procesu. Navzdory nárůstu času při zpracování odkazující stránky se tento postup vyplatí, neboť dokáže odhalit množství skrytých odkazů.

### 4.3.3 Aktualizace zpětných odkazů

Počet zpětných odkazů není kumulativní a velmi často nastává situace, kdy odkaz nebo odkazující stránka přestane existovat nebo být funkční. Z tohoto důvodu je nutná jejich zpětná kontrola. Byla testována metoda, při které se nejprve označí všechny odkazy jako smazané a následně se bude provádět jejich aktivace v případě opětovného nálezu. Tento způsob nebyl vhodný, protože při párování nově nalezených a předchozích docházelo k duplikacím. Přistoupil jsem k variantě, která sice byla nestandardní, ale zajistila funkční řešení. Postup je následující. Nejprve se smažou všechny odkazy příslušející k odkazující stránce. Poté se použije stejný algoritmus jako při hledání nového odkazu. V případě, že nebyl nalezen žádný odkaz, uloží se k této stránce informace o tomto stavu.

### 4.3.4 Optimalizace procesů

Významná část implementace byla věnována i důkladné optimalizaci všech procesů. Důraz byl kladen na znovupoužitelnost kódu, zrychlení stahovacích procesů, minimalizaci databázových přístupů a integritě uložených dat. Při testování na různých webových projektech bylo zjištěno mnoho nedostatků. Funkce bylo nutné vylepšit o celou řadu kontrolních metod a dílčích postupů, aby bylo dosaženo maximální efektivity při hledání zpětných odkazů a zároveň nedošlo ke zvýšení výpočetního času.

#### **Omezení přesměrování**

Jedním z nich bylo ověření přesměrování adresy při hledání shody s doménou. Pro regulaci časové náročnosti byla kontrola přesměrování zavedena jako post-proces, který je prováděn až po samotném hledání zpětného odkazu v seznamu externích URL. Navíc byla omezena pouze na webové stránky, na nichž nebyl nalezen ani jeden zpětný odkaz. Po testování této funkce se však ukázalo, že se tím celý proces prodloužil v průměru o polovinu času. Důvodem byl výskyt vysokého počtu interních a externích odkazů v některých katalozích. Aby složitost procesu byla alespoň z části kontrolována, zavedl jsem omezení počtu zkoumaných adres na 200 i za cenu případného nenalezení odkazu.

#### **Filtrování stránek**

Při procházení seznamu odkazujících stránek se vynechávají ty, které se již v databázi pro daný projekt nacházejí. Předchází se tím zbytečnému parsování již dříve nalezených stránek. Jelikož se při vyhledávání zpětných odkazů často procházejí i stránky, na kterých se pro danou doménu žádný odkaz nenašel, byla zavedena funkce, která je ukládá do databáze spolu s identifikátorem projektu. Seznam těchto URL adres se rovněž používá pro filtrování stránek před jejich parsováním. Tímto krokem se výrazně snížil čas při hledání zpětných odkazů.

## Přístupy do databáze

Databázová architektura byla navržena tak, aby se v ní vyskytovalo co nejméně duplicitních hodnot. Tabulka evidující odkazující stránky je vzhledem k tabulce evidující odkazy ve vztahu 1:N. Každá odkazující stránka se v databázi nachází pouze jednou a může tedy být společná pro více projektů. Toto řešení zmenšuje nejen kapacitní nároky na databázi, ale zároveň zajišťuje konzistenci dat a snižuje výpočetní čas pravidelných aktualizací. U některých tabulek byl zaveden cizí klíč (nejčastěji `cor_site_id`), který byl sice redundantní, ale díky kterému se nemuselo přistupovat k datům přes několik tabulek pomocí JOIN.

### 4.3.5 Použité technologie

V rámci vývojového týmu bylo rozhodnuto o použití PHP z důvodu snadné integrity jednotlivých modulů a nasazení do provozu. Pro vývoj jsme zvolili Nette Framework verze 2.0.1. Jedná se o populární nástroj pro tvorbu webových aplikací v PHP 5, který se rychle rozrůstá a má v ČR velkou uživatelskou komunitu, díky které je k dispozici široká škála doplňků a pluginů. Nabízí podporu pro všechny moderní technologie, kterými jsou MVC, AJAX, cool URL aj. Je poskytován s licencí GNU/GPL nebo New BSD, které umožňují použití i v komerčních projektech [15].

Aplikace je vytvářena architekturou Model-View-Controller, což zajišťuje její přehlednost a znovupoužitelnost jednotlivých segmentů. Oddělení datové a řídicí části od vzhledu umožnilo také snadnou integraci jednotlivých modulů.

Jako SŘDB používám MySQL databázi. Formát úložiště dat jsem zvolil InnoDB, protože podporuje zpracování transakcí, cizí klíče a uzamykání řádků. Pro práci s databází využívám nástroj Adminer 4.1.0. Návrh datového modelu jsem vytvářel v programu MySQL Workbench 5.2 CE, který sloužil i k vygenerování ER diagramu.

Pro vizualizaci dat z analýz jsem využil grafovou komponentu Highcharts. K zobrazení a manipulaci se záznamy z databáze mi posloužila komponenta NiftyGrid, která nabízí řádkovou editaci, hromadné operace, vyhledávání, řazení, filtrování a mnoho dalších užitečných funkcí.

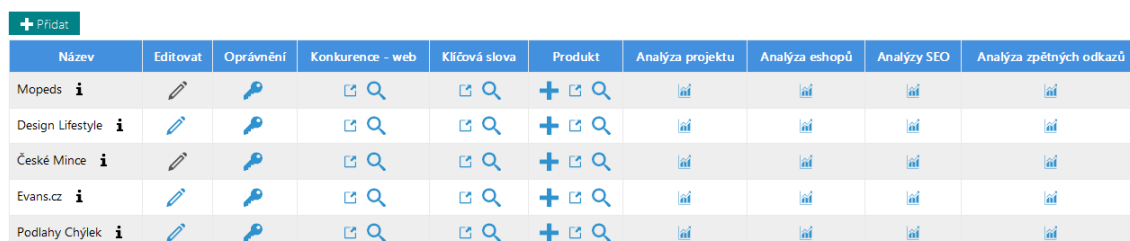
### 4.3.6 Nasazení

Testování aplikace probíhalo nejprve na lokálním počítači, kde byly laděny chyby, a probíhala optimalizace. Stahovací procesy byly spouštěny ručně a měřil se čas jejich provádění. Po dostatečném odzkoušení všech základních funkcí došlo k nasazení modulu do společného jádra aplikace na vzdáleném serveru. S použitím SVN byl modul úspěšně integrován a mohl být průběžně aktualizován po menších iteracích. Po nasazení na server byly nastaveny všechny pravidelné stahovací procesy. Pro automatizaci úloh jsem využil Cron démona, který v zadaných intervalech spouští skripty pomocí příkazu `wget`.

## 4.4 Popis funkcí systému

### 4.4.1 Správa projektu

Po registraci a přihlášení si uživatel může vytvořit nový projekt. K jednomu projektu lze přiřadit více webů, ale každý je analyzován samostatně. Weby v rámci jednoho projektu mají společná klíčová slova i konkurenty, a proto by měly být ze stejného odvětví. Po nastavení všech potřebných údajů o projektech a stažení dat je k dispozici několik sekcí, které vznikly ve spolupráci s jinými diplomanty. Uživatel se dostane ke všem analýzám z jednoho místa, viz Obrázek 13.



<a href="#">+ Přidat</a>									
Název	Editovat	Oprávnění	Konkurence - web	Klíčová slova	Produkt	Analýza projektu	Analýza eshopů	Analýzy SEO	Analýza zpětných odkazů
Mopeds									
Design Lifestyle									
České Mince									
Evans.cz									
Podlahy Chýlek									

Obrázek 13: Přehled uživatelských projektů

### 4.4.2 Analýza odkazového profilu

Samotný seznam zpětných odkazů nám nestačí k tomu, abychom získali ucelený přehled o odkazovém profilu webu. Je zapotřebí provést množství různých analýz, které poskytnou informace o jeho aktuálním stavu. Odkazující stránky se zpracovávají tak, aby se mohla určit jejich kvalita. U odkazů jsou zkoumány jejich texty, atributy a URL adresy, nad kterými se provádí analýzy.

#### Přehled vlastních webů

Na úvodní stránce projektu má uživatel k dispozici seznam všech svých projektových domén se shrnutím jejich odkazových profilů. Tabulka obsahuje celkový počet odkazů, počet odkazujících stránek, domén a IP adres, google link popularitu, Page Rank, S-rank, Alexa rank, hodnocení webu a datum posledního reportu (Obrázek 14). Kliknutím na ikonu „Zobrazit analýzu“ se dostane k podrobným informacím o odkazovém profilu webu.

Název webu	Počet odkazů	Počet odkazujících stránek	Počet odkazujících domén	Počet odkazujících IP	Google Link Popularita	Google PageRank	Seznam S-Rank	Alexa Rank	Aktualizace	Zobrazit analýzu
www.ceske-mince.cz	539	279	94	69	1	2	4	10548245	14. 4. 2014	
www.galerie-minci.cz	23	18	8	8	0	0	1	0	14. 4. 2014	

Obrázek 14: Ukázka shrnutí odkazových profilů u svých webů

## Porovnání webů s konkurencí

Okamžitě pod tabulkou se svými weby se nachází přehled všech projektových domén, včetně konkurenčních. Slouží k porovnání odkazových profilů s konkurencí (Obrázek 15).

Název webu	Počet odkazů	Počet odkazujících stránek	Počet odkazujících domén	Počet odkazujících IP	Google Link Popularita	Google PageRank	Seznam S-Rank	Alexa Rank	Hodnocení	Aktualizace	Zobrazit analýzu
<a href="http://www.zlataky.cz">www.zlataky.cz</a>	2001	277	73	53	0	3	5	271033	30.13	28. 4. 2014	<a href="#">[info]</a>
<a href="http://www.zlate-mince.cz">www.zlate-mince.cz</a>	548	225	94	68	0	3	6	815069	28.58	28. 4. 2014	<a href="#">[info]</a>
<a href="http://www.ceskamincovna.cz">www.ceskamincovna.cz</a>	391	237	69	48	0	4	5	406495	24.51	28. 4. 2014	<a href="#">[info]</a>
<a href="http://www.ceske-mince.cz">www.ceske-mince.cz</a>	539	279	94	69	1	2	4	10556285	21.34	28. 4. 2014	<a href="#">[info]</a>
<a href="http://www.e-mince.cz">www.e-mince.cz</a>	205	165	50	42	0	3	4	1093719	18.94	28. 4. 2014	<a href="#">[info]</a>
<a href="http://www.mince-numismatika.cz">www.mince-numismatika.cz</a>	242	186	34	24	0	2	4	1859779	18.26	28. 4. 2014	<a href="#">[info]</a>
<a href="http://www.mint.cz">www.mint.cz</a>	177	151	80	69	0	4	3	406495	16.41	28. 4. 2014	<a href="#">[info]</a>
<a href="http://www.numismatika-zlin.cz">www.numismatika-zlin.cz</a>	211	122	47	32	0	3	3	7528112	14.91	28. 4. 2014	<a href="#">[info]</a>
<a href="http://www.zlatoweb.cz">www.zlatoweb.cz</a>	1	1	1	1	0	0	3	0	10.75	28. 4. 2014	<a href="#">[info]</a>
<a href="http://www.minceceske.cz">www.minceceske.cz</a>	59	52	7	6	0	1	2	0	8.42	28. 4. 2014	<a href="#">[info]</a>
<a href="http://www.zlatemince.cz">www.zlatemince.cz</a>	8	7	4	4	0	0	2	0	7.32	28. 4. 2014	<a href="#">[info]</a>

Obrázek 15: Ukázka srovnání s konkurenčními weby

## Hodnocení webů

Všechny weby v tabulce jsou ohodnoceny specifickým číslem udávajícím jejich kvalitu, podle kterého jsou pak jednotlivé domény řazeny. Pro výpočet hodnocení se používá rovnice váženého průměru (

Rovnice 2) skládající se z hodnot:

- celkového počtu odkazů (PO)
- počtu odkazujících stránek (POS)
- počtu odkazujících domén (POD)
- počtu odkazujících IP (POI)
- Google Link Popularity (GLP)
- Page Rank (PR)
- S-rank (SR)

$$V = \frac{PO * 1 + POS * 3 + POD * 5 + POI * 7 + GLP * 20 + PR * 150 + SR * 100}{1 + 3 + 5 + 7 + 20 + 150 + 100}$$

Rovnice 2: Vážený průměr webu

Váhy jsou přiděleny podle důležitosti metriky. Výpočet se provádí při pravidelných reportech a jeho výsledná hodnota je ukládána.

## Přehled změn webů

Přepnutím na záložku „Poslední změny“ získá uživatel přehled posledních změn sledovaných hodnot u všech webů. V případě zvýšení je změna vizualizována šipkou směrem nahoru s příslušným číslem. Při snížení je to šipka dolů s minusovou hodnotou. Pokud ke změně nedošlo, je zobrazena pomlčka (Obrázek 16).

Název webu	Počet odkazů	Počet odkazujících stránek	Počet odkazujících domén	Počet odkazujících IP	Google Link Popularity	Google PageRank	Seznam S-Rank	Alexa Rank	Aktualizace	Zobrazit analýzu
www.designlifestyle.cz	-	-	-	-	-	-	-	-	27. 4. 2014	
www.design-lifestyle.cz	↑ 6	↑ 4	↑ 2	↑ 1	-	-	-	-	27. 4. 2014	

Obrázek 16: Ukázka posledních změn odkazového profilu u projektových webů

## Seznam odkazujících domén

Jedním z důležitých faktorů je počet odkazujících domén a jejich seznam. Uživatel přehledně vidí, jaké weby na něj odkazují a v jakém počtu. Byla zde implementována funkce na parsování URL adresy a zjištění její domény a případně subdomény. Doménový tvar se ukládá do databáze již při vkládání odkazu, aby se nemusel zjišťovat při každém zobrazení seznamu. Poté se pomocí jednoduchého dotazu získá pole všech domén a po odstranění duplicitních hodnot dostaneme seznam všech referujících webů. U každé domény je zobrazen i počet odkazů, které z ní směřují a procentuální výskyt vzhledem k celkovému počtu zpětných odkazů zkoumaného webu (Obrázek 17).

Odkazující doména ↕	Počet odkazů z domény ↕	Výskyt (%)
atlasfirem.info	13	39.4 %
firmy.cz	5	15.2 %
seonastroje.cz	2	6.1 %
dvereinterier.kvalitne.cz	2	6.1 %
najisto.centrum.cz	2	6.1 %
m.atlasfirem.info	2	6.1 %
edb.cz	1	3 %
azet.sk	1	3 %
domov-bydleni.adresarfirem.cz	1	3 %
kontaktyfirem.cz	1	3 %
edb.eu	1	3 %
firmy.adresarfirem.cz	1	3 %
katalog-webu.cz	1	3 %

Obrázek 17: Ukázka výpisu odkazujících domén

## Kvalita odkazující stránky

Dalším analyzovaným prvkem je kvalita odkazující stránky, která se posuzuje podle mnoha dílčích faktorů, mezi kterými jsou například počet odkazů na stránce, Page Rank, S-Rank, tematická podobnost a klíčová slova v textu zpětného odkazu. Kromě těchto obecně známých veličin existuje i mnoho dalších neznámých, které vyhledávací enginy nezveřejňují a kolují o nich pouze spekulace. V analýze jsem se tedy zabýval pouze měřitelnými a na pohled viditelnými



kvalitativními prvky uvedenými výše. Při budování zpětných odkazů je důležité přihlížet k tomu, jakou hodnotu nám ve výsledku přinesou. V praxi se o této hodnotě mluví jako o „link juice“. Kolik „šťávy“ přiteče do našeho webu, závisí především na velikosti ranku (Page Rank, S-Rank) a počtu interních a externích odkazů webové stránky. Do rovnice se přidává ještě tzv. damping faktor, což je pravděpodobnost, s jakou uživatel nepřeruší řetězec kliknutí až k danému odkazu (Rovnice 3). Obvykle se používá hodnota 0,85 [6]. Při výpočtu jsem vycházel z obecného vzorce pro výpočet Page Ranku [6].

$$V(L) = (1 - d) + d \frac{PR(T)}{C(T)}$$

Rovnice 3: Výpočet hodnoty odkazu

Podle tohoto výpočtu jsem stanovil kvalitu jednotlivých odkazů, která se poté používá při dalších analýzách. Weby a zejména katalogy, které mají lepší hodnocení – kvalitnější odkazy, budou přednostně doporučovány k registraci i u dalších uživatelů. Systém se takto učí ze získaných dat a využívá je ke zlepšení jiných odkazových profilů. S rostoucím počtem analyzovaných katalogů a webů bude růst i kvalita doporučování dalším uživatelům.

## Texty odkazů

Podle společnosti Google jsou texty zpětných odkazů jedním z kritérií při vyhodnocování jejich účinnosti, proto je jejich analýza velmi důležitá. Vzhledem k tomu, že na prvním místě by měla být přirozenost odkazového profilu, nesmí se na neoptimalizované texty odkazů nahlížet s despektem. Mezi tyto texty patří například „*Klikněte zde*“, „*Odkaz*“ nebo „*Zde*“ apod. Běžní uživatelé se totiž o budování zpětných odkazů nezajímají a většina z nich ani neví, co to znamená. Jedná se však o přirozené odkazy, které zde mají své místo a Google tvrdí, že jsou dokonce i žádané. Moje práce se ale zabývá optimalizací, a proto jsem se zaměřil na analýzu stávajícího stavu textů odkazů a zejména na výběr správných textů pro budování nových odkazů. V aplikaci je zobrazen seznam všech unikátních textů odkazů i s počtem jejich výskytů. Navíc je doplněn o procentuální hodnotu výskytu, vzhledem k celkovému počtu odkazů (Obrázek 18).

Text odkazu ↕	Počet výskytů ↕	Výskyt (%)
www.podlahy-chylek.cz	18	54.5 %
podlahy-chylek.cz	4	12.1 %
podlahy-chylek.cz/	4	12.1 %
http://www.podlahy-chylek.cz	3	9.1 %
podlahy martin chylek	2	6.1 %
img	1	3 %
http://podlahy-chylek.cz	1	3 %
		Záznamů na stranu: 20 <input type="button" value="Ok"/>

Obrázek 18: Ukázka seznamu textů odkazů

## Cílové stránky

Jedná se o stránky webu, na které je odkazováno. V aplikaci se zaznamenává nejenom zdrojová URL adresa, ale i cílová. V naprosté většině případů směřuje zpětný odkaz na domovskou stránku. Odkazy by však správně měly mířit na různé stránky jednoho webu tak, jak je to





zvykem u přirozeného odkazování. Tento aspekt by měl jít ruku v ruce s klíčovými slovy obsaženými v textu odkazu a s tematičností cílové stránky. V aplikaci je zobrazen seznam všech stránek webu, na který směřují zpětné odkazy, včetně četnosti výskytu.

## Seznam odkazujících stránek

Pro každou URL adresu projektu se hledají stránky, které na ni odkazují. Jeden web může odkazovat i z více svých stránek. U každé odkazující webové stránky se zjišťují hodnoty:

- URL adresa
- META title
- META description
- META keywords
- Počet interních odkazů (směřujících na stejnou doménu)
- Počet externích odkazů (směřujících na jiné domény)
- Google PageRank
- Google PageRank domovské stránky
- Seznam S-Rank
- Alexa Rank
- Link Value

Nalezené odkazující stránky jsou uživateli zobrazeny v přehledné tabulce se všemi údaji a náhledem stránky (Obrázek 19).

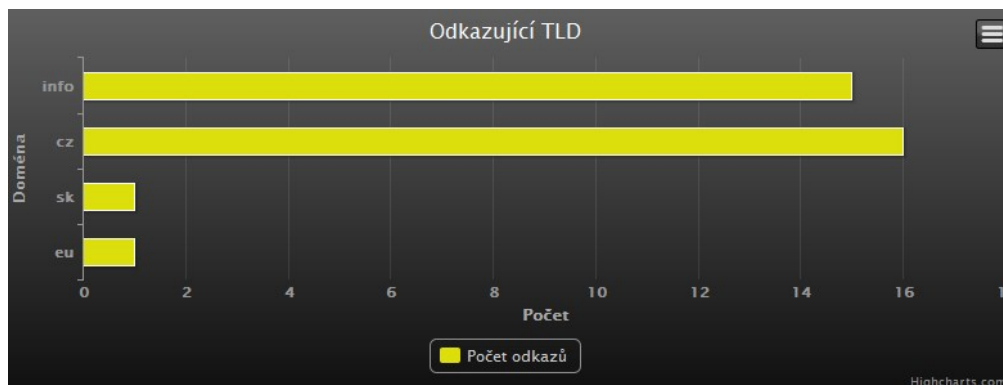
Náhled	URL adresa	Titulek	Popis	LV	IO	EO	PR	PRH	SR	AR	Stav	Akce
	<a href="http://www.dvereinterier.kvalitne.cz/">http://www.dvereinterier.kvalitne.cz/</a>	Dveře interiéru	Prodej a renovace dveří a zárubní, Rožnov,...	0.177	20	12	1	1	20	117090	ok	
	<a href="http://www.dvereinterier.kvalitne.cz/index.html">http://www.dvereinterier.kvalitne.cz/index.html</a>	Dveře interiéru	Prodej a renovace dveří a zárubní, Rožnov,...	0.177	20	12	1	1	0	117090	ok	
	<a href="http://www.seonastroje.cz/seo-monitor-ranky-vyvoj/www.podlahy-chylek.cz/11004?page=1">http://www.seonastroje.cz/seo-monitor-ranky-vyvoj/www.podlahy-chylek.cz/11004?page=1</a>	Vývoj ranků webu www.podlahy-chylek.cz	Vývoj ranků webu www.podlahy-chylek.cz	0	41	14	-	5	0	263427	ok	
	<a href="http://www.seonastroje.cz/seo-monitor-ranky-vyvoj/www.podlahy-chylek.cz/11004?page=2">http://www.seonastroje.cz/seo-monitor-ranky-vyvoj/www.podlahy-chylek.cz/11004?page=2</a>	Vývoj ranků webu www.podlahy-chylek.cz	Vývoj ranků webu www.podlahy-chylek.cz	0	41	14	-	5	0	263427	ok	

Obrázek 19: Ukázka výpisu odkazujících stránek

Databáze odkazujících stránek je společná pro všechny weby a projekty, aby nedocházelo ke zbytečnému parsování stránek, které již v systému existují.

## Domény nejvyšší úrovně (TLD)

Pro určení míry různorodosti odkazového profilu je důležitým aspektem i analýza domén nejvyšší úrovně, tzv. TLD. Uživatel tak získá přehled o tom, z jakých národních nebo generických domén na něj směřují odkazy. Pokud je web propagován i v zahraničí, může tato analýza poskytovat informace sloužící k lepší specifikaci oblastí pro budování zpětných odkazů. Na různých typech TLD by však nemělo záviset celkové hodnocení webu, a proto se u nich neprovádí žádná optimalizace. K vyobrazení počtu jednotlivých domén je použit sloupcový graf (Obrázek 20).



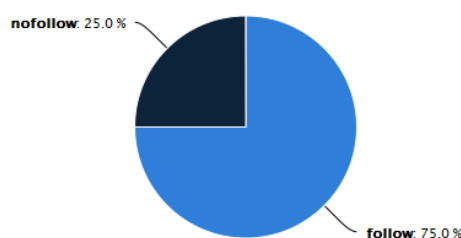
Obrázek 20: Ukázka grafu odkazujících TLD

## Google Link Popularity

Jedná se o metriku udávající počet odkazujících stránek, které se objevují ve výsledcích vyhledávání při použití operátoru „*link:*“ ve vyhledávači Google. Aplikace si při stahování označuje tyto odkazy speciálním indexem, aby byly později snadno k vyhledání. Tyto „Google“ odkazy by měly vypovídat o jakési vyšší kvalitě nebo autoritě oproti ostatním, jinak by ve výsledcích zobrazeny nebyly. Počet je uvedený v pravidelném reportu a v souhrnu analýzy.

## Graf typů odkazů

Pro určení kvality odkazu je důležité znát, jestli má povoleno přenášet svou hodnotu. Některé weby přidávají k odkazu atribut `rel="nofollow"`, což jej činí „bezcným“. K získání lepšího povědomí o odkazech ve svém profilu je k dispozici koláčový graf se všemi typy odkazů (Obrázek 21).



Obrázek 21: Ukázka grafu typů odkazů

## Seznam důležitých domén

Na českém a světovém internetu jsou weby a domény, které mají vysokou autoritu, zažitou historii a dobré hodnocení od vyhledávačů. Existence těchto domén ve svém odkazovém profilu je téměř nezbytná. Nemusí přinášet přímo kvalitní odkazy, ale mohou být dobrým zdrojem návštěvnosti. Jedná se například o velké firemní a webové katalogy, agregátory zboží nebo prodejní portály. Vybral jsem seznam důležitých domén a uživateli jej zobrazuji ve formě kontrolní tabulky (Obrázek 22).

Odkazy z důležitých domén	
Zboží.cz	✓
Firmy.cz	✓
Centrum.cz	✗
Adresarfirem.cz	✗
Wikipedia.org	✓
Aukro.cz	✓
Heureka.cz	✓

Obrázek 22: Ukázka kontroly důležitých domén

## Srovnání s konkurencí

U komerčních webových stránek je velmi důležité i srovnání s konkurencí. Uživatel má možnost přiřadit k vlastnímu projektu i konkurenční URL adresy, pro které se rovněž provádí analýza odkazového profilu. V systému se pak tato data zobrazují ve formě srovnávací tabulky spolu s vlastním projektem. Majitel stránek tak vidí, ve kterých oblastech je na tom lépe nebo naopak hůře oproti konkurenci. Stejnou analýzu jako pro svůj web si může zobrazit i pro konkurenci, aby mohl podrobně zkoumat její data a poté je použít jako předlohu pro vlastní budování.

## Odkazy na sociálních sítích

V poslední době se rozrůstá trend propagování webových stránek a firemních profilů na sociálních sítích. S tím je spojeno i budování zpětných odkazů v této oblasti a je za potřebí jej sledovat. Proto v odkazovém profilu webu zobrazuji i četnost výskytu jeho URL adresy na jednotlivých sociálních sítích. Zaměřil jsem se na servery:

- Facebook (počet „Líbí se mi“, počet sdílení a počet komentářů)
- Google Plus (počet +1)
- LinkedIn
- Twitter

## Rank webu

Celý proces optimalizace webu a budování zpětných odkazů je měřitelný pomocí různých metrik. Při pravidelných reportech se zjišťují a zaznamenávají hodnocení webu, kterými jsou:

- Google Page Rank
- Seznam S-rank
- Alexa Rank

Díky jejich pravidelnému evidování může uživatel sledovat vývoj webu a ověřit si, zda jsou jeho postupy správné a přínosné. Data jsou zobrazena jak formou tabulky, tak i spojnícovým grafem s časovou osou. Hodnoty jsou aktualizovány jednou týdně.

### 4.4.3 Budování zpětných odkazů

Kromě analýzy stávajícího stavu odkazového profilu webu nabízí aplikace i nástroj k budování zpětných odkazů. Uživatel má možnost vybrat si ze stovek domácích i zahraničních katalogů, kam může umístit reference na svůj web.

#### Vytvoření budovacího plánu

Prvním krokem celého procesu je vytvoření plánu budování odkazů. Uživatel uvádí jeho název, popis a vybere oblast registrace:

- Obecné katalogy
- PR katalogy
- Tematicky podobné weby

#### Tvorba obsahu

Systém má za úkol evidovat textová data použitá při registraci do vybraných katalogů, aby byla zajištěna co největší různorodost obsahu a rovnoměrnost jejich výskytu. Před samotným procesem výběru vhodných webů je zapotřebí vyplnit registrační údaje o firmě:

- Název společnosti
- Kontaktní osoba
- IČ
- DIČ
- Adresa (ulice, č. p., město, PSČ, kraj)
- Telefonní číslo
- E-mail













Dále je nutné vložit do systému několik verzí textu a popisů, které jsou potřebné k registracím. Evidovanými údaji jsou:

- Titulek, nadpis
- Krátký popis webu nebo činnosti firmy
- Dlouhý popis (článek)
- URL adresa odkazu

Obsahová registrační data jsou automaticky přiřazena k vybranému katalogu. Pokud se jedná o obecný nebo firemní katalog, použijí se data s krátkým popisem. V případě registrace do PR katalogu bude použit záznam obsahující dlouhý popis (článek). Systém zároveň dbá na rovnoměrné použití textů, a proto vždy vybírá záznam s nejmenším počtem použití a poté u něj inkrementuje jeho hodnotu o 1. Tím dochází ke snížení duplicity textů na internetu, což může pomoci k vylepšení hodnocení webu.

## Seznam katalogů v plánu

Ještě předtím, než si uživatel vytvoří budovací plán, se mohou ukládat katalogy, v nichž byly odkazy nalezeny při jejich vyhledávání. Katalogy tohoto typu nemají přiřazena žádná registrační data a jsou ve stavu „publikováno“. Po vytvoření a otevření plánu má uživatel k dispozici již seznam odkazujících katalogů (Obrázek 23). Seznam webů „k výběru“ a „v plánu“ je tak aktuální ještě před samotným začátkem budování.

<input type="checkbox"/> Nahled	URL adresa katalogu	PR	SR	Prům. LV	Kategorie	Jazyk	Cena	Stav	Datum přidání	Datum poslední kontroly	Data	Akce
	<input type="text"/>				Katalogy firm	--	--					Filtrovat
	http://sluzby.cz	5	5	--	PR články	Český	Zdarma	Registrováno	2014-03-29 11:51:51			
	http://hyperinzerce.cz/	0	9	0.154	Nefunkční	Český	Zdarma	Přidáno	2014-03-29 11:26:16			 
	http://kontaktyfirem.cz	3	3	0	Specializované	Český	Zdarma	Publikováno	2014-03-10 15:06:06	2014-03-31 14:13:19		
	http://adresarfirem.cz	4	5	0.046	PR články	Český	Zdarma	Registrováno	2014-03-10 11:18:53			

Obrázek 23: Seznam katalogů v plánu

Kvůli kontrole průběhu registrací na jednotlivých webech se každý katalog umístěný „v plánu“ nachází v určitém stavu. Ihned po přidání je ve stavu 0, který jako jediný dovoluje odstranění. Další stavy jsou následující:

- 1 – posláno a uloženo do desktopové aplikace (přes API)
- 2 – posláno do katalogu k registraci
- 3 – registrováno (po ověření e-mailem)
- 4 – publikováno (nalezeno aplikací)
- 5 – nedokončeno (nezdařená registrace)
- 6 – odstraněno (nenalezeno přes vyhledávač)

### **Výběr katalogů**

Po kliknutí na předem vytvořený budovací plán se uživateli zobrazí seznam webů vhodných k registraci. Katalogy a weby jsou doplněny o informace, podle kterých si uživatel může vybrat vhodnou variantu. K dispozici je řazení a filtrování podle:

- Ranku (Page Rank, S-rank, Alexa Rank)
- Průměrné link value (výpočet viz 4.4.4)
- Kategorie

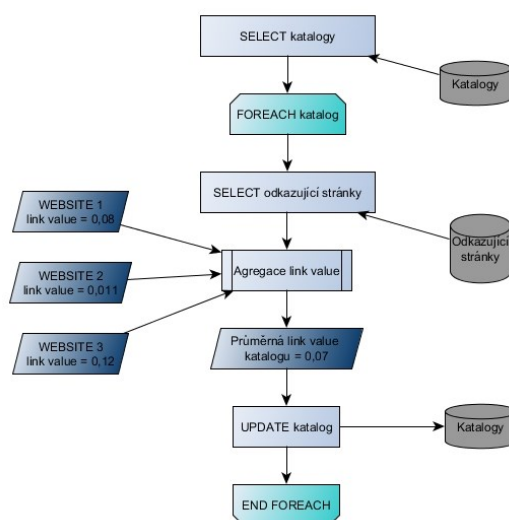
Pokud má uživatel řádně vyplněna registrační data, může přidat katalog do plánu a provést jeho registraci ručně (rychlejší a spolehlivější) nebo to nechat na automatizaci (v případě používání desktopové aplikace). Vybraný web je přesunut ze seznamu „k výběru“ do seznamu webů „v plánu“. K dispozici je i tlačítko na automatický výběr 10 katalogů podle zvolené metriky. Uživatel má na výběr z 3 kritérií - průměrná link value, Page Ranku nebo S-rank.

### **4.4.4 Využití databáze znalostí**

Zjišťováním odkazových profilů různých webových projektů se vytváří a rozšiřuje databáze znalostí o nalezených odkazujících stránkách a jejich přínosu z hlediska hodnoty zpětného odkazu. Tyto informace mají využití pro budování nových odkazů. Na základě vypočtené hodnoty, kterou dokáže stránka skrze odkaz předat, ji lze doporučit dalším uživatelům. Registrační katalogy si uživatelé primárně vybírají podle ranku jejich domovské stránky, což ale nevypovídá o jejich skutečné kvalitě. Vložené odkazy spolu s popisem totiž nikdy nejsou umístěny přímo na hlavní straně, ale někde uvnitř webu, kde jsou ranky podstatně nižší.

### **Výpočet průměrné hodnoty odkazu z katalogu**

Při naplňování databáze odkazujícími stránkami dochází k průběžnému výpočtu hodnot jejich odkazů. Následně se pro každý, z téměř 3000 katalogů, naleznou odpovídající odkazující stránky, a provede se aritmetický průměr hodnot jejich odkazů. Celý proces je znázorněn vývojovým diagramem (Obrázek 24). Na základě takto vypočtené hodnoty je možné lépe určit a doporučit katalogy, které skutečně mohou přinést skrz zpětný odkaz nějakou „link juice“.



Obrázek 24: Diagram výpočtu hodnoty katalogu

## Doporučování odkazů konkurence

Dalším využitím stažených dat je doporučování webů, na které odkazuje konkurence. Pokud má konkurenční web lepší hodnocení, může jeho odkazový profil sloužit jako návod pro budování vlastních odkazů. Data by měla sloužit pro rychlejší vyhledání vhodných stránek podle jejich skutečné, již vypočtené hodnoty, a vybrat z nich ty, které nejlépe vylepší jeho profil.

### 4.4.5 Vyhledávač tematických webů

Pro účely budování nových zpětných odkazů jsem implementoval funkci na vyhledávání a doporučování tematických webů. Na rozdíl od registrací do katalogů, kde je předem připraven seznam webů, se zde provádí jejich vyhledání pomocí předem zadaných projektových klíčových slov. Pro každé klíčové slovo se nalezne 0-100 výsledků, které se ukládají do databáze. Pro předložení kvalitnějšího seznamu webů je aplikován filtr, který ignoruje vlastní i konkurenční domény. Aby měl uživatel představu o kvalitě a relevantnosti nabídnutých stránek, stahuje o nich systém nejrozumnější informace. Podle těchto dat si může vybrat, na kterou z nich by bylo vhodné umístit odkaz. Zjišťované a evidované metriky webu jsou:

- Titulek
- Popis
- Page Rank
- S-rank
- Počet odkazů na stránce (interních i externích)
- Link Value

Kromě klasických webů se vyhledávají i blogy a diskuzní fóra. V seznamu pak uživatel vidí, o jaký typ se jedná. Stejně jako u katalogů a PR webů je i zde možnost přidání tematické stránky do budovacího plánu (Obrázek 25).



Náhled	URL	Titulek	Popis	PR	SR	Počet odkazů	Link Value	Typ	Akce
	<a href="http://femme-mince.blog.cz/">http://femme-mince.blog.cz/</a>	Femme Mince	Krásný den! Dnes mám pro vás jeden z mých nejrychlejších receptů na koláč. Je to pochoutka na způsob tvarožníku, provoněná vanilkou a... Blog.cz - Stačí otevřít a budeš v obraze.	2	30	59	0.179	blog	+ x
	<a href="http://halamickova.blog.idnes.cz/c/364382/Šmejdi-aneb-mince-o-dvou-stranach.html">http://halamickova.blog.idnes.cz/c/364382/Šmejdi-aneb-mince-o-dvou-stranach.html</a>	Šmejdi aneb mince o dvou stranách. Blog - Petra Halamičková (blog.IDNES.cz)	Dlouho jsem se těšila, až budu v TV dávat Šmejdy, protože se o nich mluvím už od doby, kdy byli natočeni. Za tuto práci, odvalu a nasazení patří slečně režisérce velký dík. Mám kolem sebe mnoho lidí ve věkovém rozmezí 65-89 let, a tak o různé historky z p	0	20	180	0	blog	+ x
	<a href="http://coins.blog.cz/0808/cisteni-minci">http://coins.blog.cz/0808/cisteni-minci</a>	Čištění mincí   coins.blog.cz	Mnoho z došlých dotazů na které odpovídám, se mimo jiné týkají čištění mincí. Jak správně vyčistit mince? Podívejme se společně na... Blog.cz - Stačí otevřít a budeš v obraze.	-1	30	40	0	blog	+ x
	<a href="http://moucka.blog.idnes.cz/c/360720/Putin-a-Obama-dve-strany-teze-mince.html">http://moucka.blog.idnes.cz/c/360720/Putin-a-Obama-dve-strany-teze-mince.html</a>	Putin a Obama: dvě strany téže mince. Blog - Jaroslav Moučka (blog.IDNES.cz)	Globální šach finanční oligarchie se stupňuje. Dlouho jsem přemýšlel, jestli kroky, které prezident Putin posledních několik let podnikal jsou skutečnou opozicí proti zájmům této zločinecké kliky. Poslední vývoj okolo Sýrie společně se zraněním vedoucího	0	10	183	0	blog	+ x

Obrázek 25: Ukázka výběru tematických webů

## 4.4.6 Automatizace procesů

Většina procesů není kvůli své časové náročnosti prováděna ihned po zadání uživatelem, ale je spouštěna automaticky v různých časových intervalech. K tomuto účelu se používají Cron úlohy, které jsou nastaveny na určitý čas a periodu a vykonávají posloupnost zadaných skriptů.

### Stahovací procesy

- Vyhledávání pomocí SEOmoz API
- Vyhledávání na Google podle URL adresy
- Nalezení Google Link Popularity
- Prohledání seznamu katalogů
- Hledání na důležitých doménách:
  - firmy.cz
  - centrum.cz
  - zboží.cz
  - aukro.cz
  - heureka.cz
  - adresarfirem.cz

### Aktualizační procesy

- Kontrola odkazujících stránek
- Pravidelné reporty

## 4.4.7 Administrační rozhraní

Pro potřeby správce systému bylo vytvořeno administrační rozhraní, které umožňuje manuální spuštění libovolného stahovacího procesu. Na vyžádání může být spuštěna jakákoliv úloha, na kterou by se jinak muselo čekat. Všechny důležité procesy běží na pozadí a pravidelně

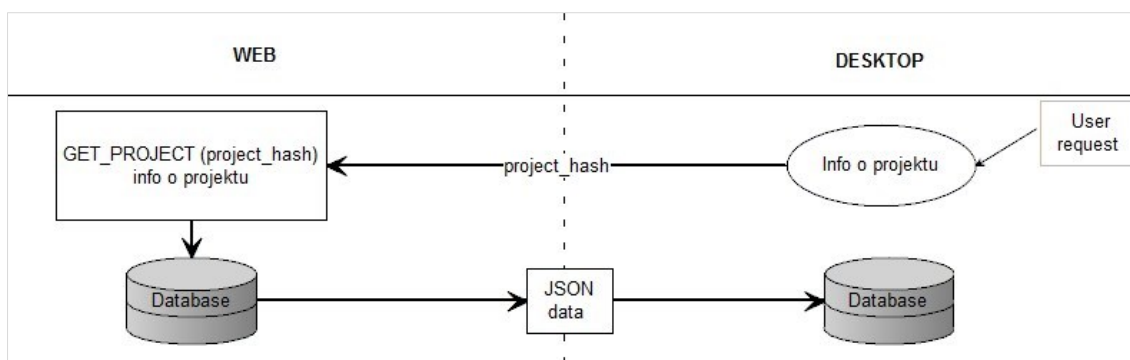
se spouští v nastavených intervalech. Zákazník nebo správce si může vyžádat urychlení stahování nebo přepočítání dat, k čemuž slouží právě tato funkce.

#### 4.4.8 API

V rámci tohoto projektu bylo vytvořeno API rozhraní pro komunikaci a výměnu dat mezi webovou a desktopovou aplikací, která zajišťovala automatickou registraci do internetových katalogů a PR webů. Aplikaci vyvíjel Jiří Kopel v rámci své bakalářské práce. Serverová část poskytuje databázi pro uložení informací o uživatelských projektech, seznamu katalogů a seznamu doporučených katalogů k registraci. Rozhraní komunikuje oběma směry, proto nabízí možnost exportu i importu dat v prostředí serverové databáze.

##### Export dat o projektu

Funkce je určena pro uživatele, kteří již využívají webový analytický nástroj, ve kterém mají vyplněny všechny potřebné údaje. Po instalaci desktopové aplikace pro registraci do katalogů si mohou jednoduše importovat informace o svém projektu do lokální databáze. Pro ověření vlastnických práv budou data poskytnuta pouze po porovnání HASH kódu projektu, který uživatel vyplní v desktopové aplikaci, a který je shodný s kódem v globální databázi. Komunikace je znázorněna diagramem níže (Obrázek 26).



Obrázek 26: API export dat o projektu

Přenášená data o projektu:

- Název projektu
- Popis projektu
- Datum vytvoření

Další přenášená data již souvisí se samotným registrováním do katalogů a jedná se o seznam URL adres přidružených k projektu:

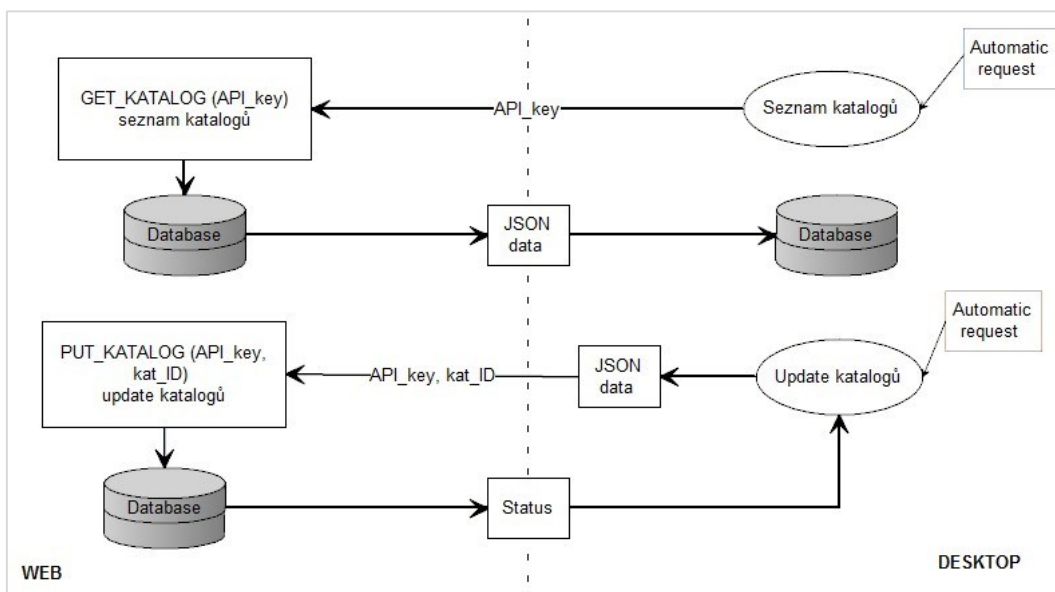
- URL adresa
- ID URL adresy

Registrační informace o společnosti:

- Název firmy
- Kontaktní jméno
- IČ
- DIČ
- Adresa firmy (ulice, město, PSČ, okres)
- Telefonní číslo
- E-mailová adresa

## Export / Import seznamu katalogů

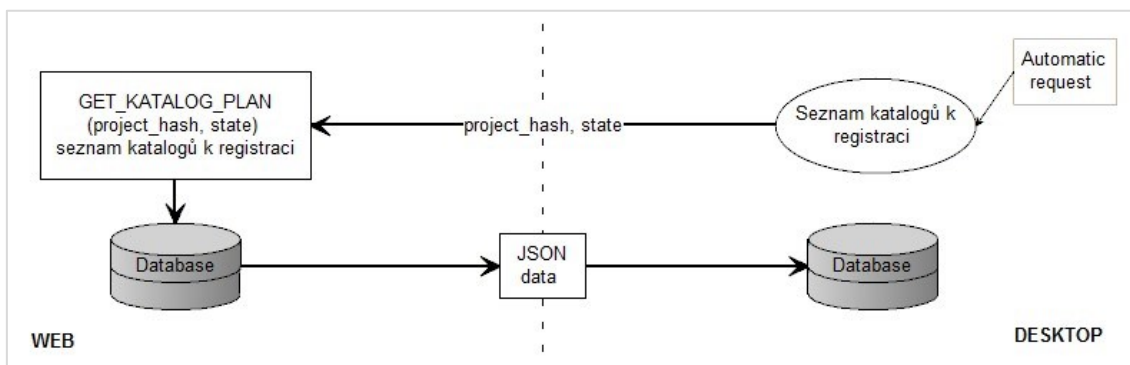
V serverové databázi je uložen seznam katalogů a PR webů sloužících k budování zpětných odkazů. Desktopová aplikace využívá tuto databázi k uložení dat o jednotlivých katalogích, která se ukládají ve formátu XML a jsou získávána při registracích. Obsahují informace o formulářových prvcích vyplňovaných v průběhu registrace, které se zaznamenávají pro další použití. Globální databáze zajišťuje přístup k těmto datům i pro nové uživatele systému. Popis procesu je znázorněn diagramem uvedeným níže (Obrázek 27).



Obrázek 27: API - import/export katalogů

## Export katalogů k registraci

Aplikace si bude stahovat seznam katalogů, do kterých chce uživatel provést registraci. Spolu s URL adresami projektových webů a katalogů budou poskytnuty i optimalizovaná registrační data jako je text odkazu, popis vloženého webu a v případě PR webů i články. Přenos informací je zobrazen diagramem viz Obrázek 28.



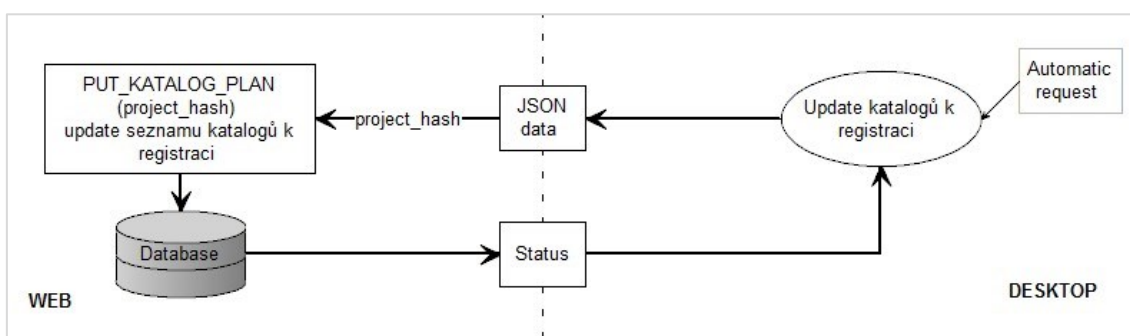
Obrázek 28: API - export katalogů k registraci

## Aktualizace seznamu katalogů k registraci

Desktopová aplikace bude u katalogů k registraci měnit jejich stav a datum provedení. Stavy registrací jsou:

- 0 – přidáno k registraci
- 1 – posláno a uloženo do desktopové aplikace
- 2 – posláno do katalogu k registraci
- 3 – registrováno (po ověření e-mailem)
- 4 – publikováno (nalezeno přes vyhledávač)
- 5 – nedokončeno (nezdařená automatická registrace)
- 6 – odstraněno (nenalezeno přes vyhledávač)

Aktualizace stavů je popsána diagramem níže (Obrázek 29)



Obrázek 29: API - update seznamu katalogů k registraci

## Formát a zabezpečení dat

Data jsou přenášena jako asociativní pole ve formátu JSON. Klíče a ID hodnot z databáze jsou přenášeny pomocí metody GET. Veškerá data posílána do webové databáze jsou v metodě POST.

## 5 Modul pro stahování aukcí z portálu Aukro.cz

Aukro.cz je v současnosti nejpoužívanějším a nejznámějším aukčním portálem v ČR a je stále více využíván obchodníky k prodeji svého sortimentu. Zboží je čím dál více nabízeno jako nové se zárukou a ve stejné kvalitě jako v klasickém e-shopu. Obchodník, který chce aktivně sledovat konkurenci a reagovat na její cenovou politiku a strategii, musí vzít do úvahy i tento server. Proto v rámci společného analytického nástroje vznikl modul, který stahuje ukončené aukce z portálu Aukro.cz. Slouží k vyhledání aukcí podle klíčových slov produktů a k jejich evidenci. Získaná data budou použita pro analýzu prodeje jednotlivých produktů.

Každý projekt v systému má importován seznam svých produktů a u každého z nich jsou uvedena klíčová slova, která se používají pro vyhledání stejného produktu na Aukru.

### 5.1 Popis funkcí modulu

#### 5.1.1 Zobrazení aukcí u produktu

U každého produktu v systému, který má vyplněna klíčová slova, lze sledovat jeho nabídku a prodej na Aukro.cz. Evidují se pouze ukončené aukce, nikoliv probíhající. Po úspěšném stažení dat má uživatel možnost si zobrazit seznam aukcí, v nichž se vyskytuje hledaný produkt (Obrázek 30). Evidovanými údaji jsou zde:

- Název aukce
- Vytvářecí cena
- Konečná cena
- Počet příhozů
- Prodán (Ano/Ne)
- Kategorie
- Prodejce
- Datum ukončení aukce

Název aukce	Vytvářecí cena	Konečná cena	Prodáno	Příhozy	Kategorie	Prodejce	Konec aukce
200 Kč 1997 Emauzy - PROOF mince i certifikát!	1	2750	ANO	10	Sběratelství	7521662	2014-01-11 15:50:00
200 Kč 1997 Emauzy - PROOF mince i certifikát!	1	2827	ANO	13	Sběratelství	7521662	2014-02-23 17:55:00
200 Kč 1997 Emauzy - PROOF mince i certifikát!	2990	3040	ANO	2	Sběratelství	35930220	2014-03-16 21:23:28
200 Kč 1997 Emauzy - PROOF mince i certifikát!	2990	2990	ANO	0	Sběratelství	35930220	2014-03-24 08:13:48
Ag. mince 200 Kč Klášter EMAUZY Proof, 1997.	1	3550	ANO	8	Sběratelství	6277354	2014-04-06 15:45:00

Záznamů na stranu: 20

Obrázek 30: Seznam ukončených aukcí produktu

Název aukce je ve formě HTML odkazu, který umožňuje přímý proklik na stránku s ukončenou aukcí. Stejným způsobem lze zobrazit i stránku s prodejcem a zjistit tak jeho hodnocení a případně další nabídky.

Pod seznamem aukcí je zobrazen seznam vložených nebo generovaných frází, které se používají k vyhledávání aukcí pro daný produkt (Obrázek 31). Fráze se generují automaticky ze seznamu klíčových slov, ale uživatel je může přidávat i ručně. K dispozici je i využití tzv. nepovolených slov, která slouží k vytřížení nechtěných výsledků. Pro lepší specifikaci produktu se doporučuje uvádět u frází i cenové rozmezí. Záznamy je možné editovat, deaktivovat tlačítkem play/stop nebo úplně smazat. Pozastavená fráze se poté nepoužívá při vyhledávání. U každé fráze je možnost zobrazení pouze jejích výsledků.

Fráze	Nepovolená slova	Cena od	Cena do	Aktualizace	Typ vložení	Stav	Akce
<input type="checkbox"/> mince klášter na slovanech proof		2400	9000	2014-04-08 19:44:51	Generovaná	Aktivní	    
<input type="checkbox"/> mince klášter na slovanech špičková kvalita		2400	9000	2014-04-08 19:44:54	Generovaná	Aktivní	    
<input type="checkbox"/> stříbrná mince klášter na slovanech proof		2400	9000	2014-04-08 19:44:57	Generovaná	Aktivní	    
<input type="checkbox"/> stříbrná mince klášter na slovanech špičková kvalita		2400	9000	2014-04-08 19:44:59	Generovaná	Aktivní	    
<input type="checkbox"/> mince Emauzy proof		2400	9000	2014-04-08 19:45:03	Generovaná	Aktivní	    
<input type="checkbox"/> mince Emauzy špičková kvalita		2400	9000	2014-04-08 19:45:06	Generovaná	Aktivní	    
<input type="checkbox"/> stříbrná mince Emauzy proof		2400	9000	2014-04-08 19:45:09	Generovaná	Aktivní	    
<input type="checkbox"/> stříbrná mince Emauzy špičková kvalita		2400	9000	2014-04-08 19:45:11	Generovaná	Aktivní	    
Označené: Aktualizovat <input type="button" value="Potvrdit"/>							Záznamů na stranu: 20 <input type="button" value="Ok"/>

Obrázek 31: Vyhledávací fráze produktu pro Aukro.cz

## 5.2 Popis implementace

### 5.2.1 Proces vyhledání aukce

K vyhledání aukce se používají klíčová slova uvedená u produktu v systému. Ty se dělí na 4 skupiny:

- prefix
- povolená klíčová slova
- sufix
- zakázaná slova

Byla implementována metoda, která z těchto skupin vygeneruje vyhledávací fráze v pořadí *<prefix>* *<povolené slovo>* *<sufix>* a uloží je do databáze. Tento postup se používá k nalezení co největšího množství relevantních produktů. Zakázaná slova se využívají až po nalezení výsledku, a to k odfiltrování nesouvisejících aukcí.

Data se stahují pouze z aukcí, které již skončily, nikoliv z probíhajících. Evidují se jak prodané, tak i neprodané produkty, kvůli zkoumání celkového počtu nabídek.

Pro stahování nabídek jsem se rozhodl využít služby Aukro WebAPI, která nabízí dostatečné množství dat a nastavení potřebných pro účely tohoto modulu.

## 5.2.2 Aukro WebAPI

Portál Aukro.cz poskytuje vlastní API pro sledování a stahování nabídek, vystavování zboží a zjišťování informací o aukcích. Funguje na základě protokolu SOAP a pomocí technologie RPC zasílá XML zprávy klientovi [13]. Služba je zdarma, ale pro její používání je nutné mít platný účet a zažádat si o API klíč. Počet stahování není nijak limitován, avšak maximální počet výsledků jednoho dotazu je omezen na 100. Rozhraní nabízí metodu *doSearch*, která umožňuje vyhledání nabídek podle zvolených kritérií [13]. Výhodou je, že aukce se uchovávají až po dobu 3 měsíců, takže uživatel dostane hodnotné informace již po prvním stažení.

## 5.2.3 Automatizace procesů

Procesy, které stahují nebo aktualizují data, jsou spouštěny pomocí Cron úloh automaticky v nastavených intervalech. Pro spuštění skriptu jsem využil příkaz *wget*.

### **Stahování aukcí**

Stahování aukcí se provádí 1x denně v počtu 1000 frází. Cyklus u frází je nastaven tak, že se načítají pouze ty, jejichž datum poslední aktualizace je starší než 7 dní.

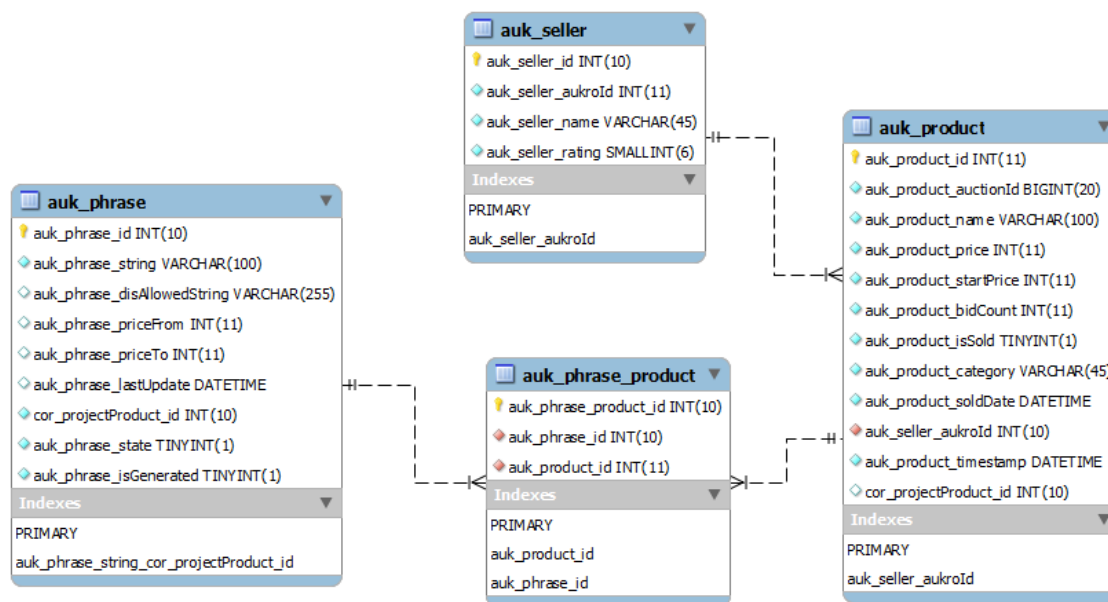
### **Generování frází**

Jednou denně se provádí generování vyhledávacích frází z klíčových slov. Proces funguje tak, že již existující fráze se znovu nekládají, ale do databáze se přidávají pouze nové. Jelikož proces není příliš časově náročný, spouští se pro všechna klíčová slova u všech projektů.

## 5.3 Databázový model

Tabulky evidují skončené aukce, vyhledávací fráze a prodejce. Nalezené produkty (výsledky aukcí) jsou s frází propojeny přes vazební tabulku (Obrázek 32). Jelikož se prodejci často opakuji, je pro ně zřízena samostatná tabulka, aby se zamezilo zbytečné redundanci dat.

### 5.3.1 ER-diagram



Obrázek 32: ER-diagram databáze Aukro modulu



# 6 Zhodnocení dosažených výsledků

## 6.1 Testování on-page SEO technik

Otestoval jsem účinnost různých SEO technik na webu, který byl špatně hodnocen vyhledávači a vůbec se neobjevoval ve výsledcích vyhledávání na potřebná klíčová slova. Neobjevoval se dokonce ani po zadání názvu webu. Na tomto stavu měl velký podíl hlavně fakt, že web nebyl v minulosti nijak optimalizován. Aplikoval jsem znalosti získané studiem této problematiky, abych otestoval jejich účinnost a dosáhl lepších výsledků.

### **Meta titulek**

Na první pohled bylo vidět, že jednotlivé webové stránky nemají správně vyplněný titulek stránky. Neobsahoval potřebná klíčová slova ani název značky. Po jejich optimalizaci se hlavní stránka začala do dvou dnů zobrazovat v SERP u předních vyhledávačů. Stačilo pouze zadat primární a sekundární klíčová slova a název webu.

Po tomto úspěšném zjištění bylo zapotřebí vytvořit unikátní titulky pro každou stránku webu. Opět po dvou dnech bylo znát zlepšení a každá stránka webu byla samostatně k nalezení v SERP. Jelikož se jednalo o lokální firmu, tak mě napadlo doplnit titulky o název města, což z pohledu vyhledávače posunulo stránku o pár míst v případě uvedení názvu města v dotazu. Při uvádění lokalizace v titulku by se měl stejný název objevovat i v samotném textu stránky, aby mu vyhledávač přidal větší důležitost.

### **Meta popis**

Stránky se konečně zobrazovaly do té samé pozice ve výsledcích, ale chyběl u nich META popis, který by lépe specifikoval obsah stránky. Popis neslouží ke zlepšení pozic, ale spíše k upoutání zákazníka a zvýšení prokliku. Napsal jsem jej tedy jako reklamu, která obsahovala klíčová slova. Výsledky sice nebyly měřitelné, ale na pohled vypadala položka v seznamu lépe než nic, protože lákala svým textem s tučně označenými relevantními slovy.

### **Textový obsah**

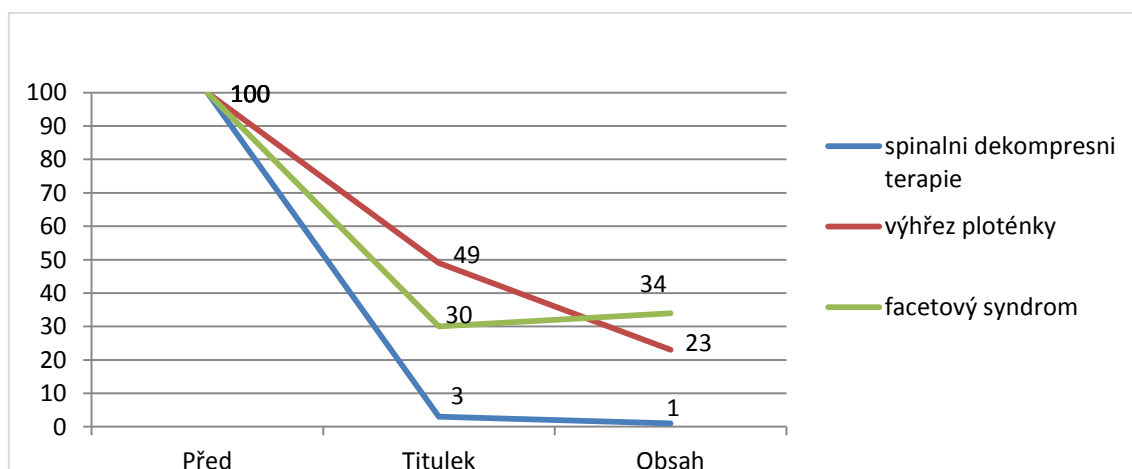
Jelikož byl obsah stránek dobře strukturovaný a originální, nebylo potřeba jej příliš měnit. Doplnil jsem pouze pár klíčových slov kvůli jejich správné hustotě. Provedl jsem také doplnění klíčových frází do různých prvků stránky, například do nadpisů H1, H2, „alt“ atributů obrázků a navigačního menu, aby byly relevantní jak k obsahu, tak i k titulku.

## Indexování stránek

Prověřil jsem také indexaci všech stránek webu ve vyhledávacích prostřednictvím příkazu *site:*. Zjistil jsem, že některé z nich nejsou vůbec indexovány, a proto se nemohly zobrazit ve vyhledávání. Vyhledávače poskytují možnost prostřednictvím formuláře vložit svůj web nebo jeho část k indexování. U Google se změna projevila do jednoho, u Seznamu zhruba do dvou dnů.

## Změna pozic

Před procesem optimalizace byly změřeny pozice hlavních klíčových slov ve vyhledávacích Seznam a Google. Na obrázku (Obrázek 33) je vidět graf změn pozic u 3 hlavních klíčových slov v jednotlivých etapách.



Obrázek 33: Graf vývoje pozic klíčových slov v Google

Z grafu je zřejmé, že největší podíl na zlepšení pozic měla změna titulku stránky (Obrázek 33). Další změny byly už pouze symbolické a nedá se jim klást velký význam, neboť ke změnám dochází i bez cílené optimalizace.

Tímto testováním vyšlo najevo, že správný titulek stránky je velmi důležitý a razantně ovlivňuje pozice ve vyhledávacích. Změny se projevily ihned po aktualizaci stránky v indexu, a to do 3 dnů od úprav a poslání URL adresy vyhledávači.

## 6.2 Analýza odkazového profilu

Aplikaci jsem otestoval na několika projektech, ke kterým jsem měl přístup a povolení majitele ke zveřejnění jejich analýz. Weby jsem volil tak, aby byly navzájem různorodé a test byl objektivní. Do testování jsem zahrnul jak web, jehož majitel již prováděl optimalizaci a budoval zpětné odkazy, tak i mladší web, který má jen minimum zpětných odkazů.

## 6.2.1 Porovnání výsledků s konkurenčními řešeními

V rámci testování aplikace jsem porovnal své výsledky s dostupnými konkurenčními nástroji, abych zjistil rozdíly v počtu nalezených zpětných odkazů a jiných parametrů.

Velké konkurenční nástroje mají výhodu v tom, že využívají databáze indexovaných webů získaných procházením jednotlivých stránek internetu, a tak mají v relativně krátkém čase přehled o zpětných odkazech zkoumaného webu.

Relevantní metrikou pro porovnání je počet odkazujících domén a stránek. Z tabulek je patrné, že rozdíly v celkovém počtu odkazů jsou odlišné, neboť každý nástroj má jinou metodiku zobrazování výsledků (Tabulka 5). Někteří evidují u každé odkazující stránky pouze jeden odkaz. V mém řešení se evidují všechny zpětné odkazy z referujících stránek.

### České-mince.cz

nástroj	Web CEO	GWT	Seo SpyGlass	moje řešení
celkový počet odkazů	570	10282	0	539
počet odkazujících stránek	491	-	662	279
počet odkazujících domén	170	170	205	94
počet odkazujících IP	115	-	92	69
Google Link Popularity	1	-	-	1
počet anchor textů	7	-	31	121
počet různých TLD	5	-	7	7
počet cílových URL	2	100	9	161

Tabulka 5: Srovnání výsledků pro web ceske-mince.cz

U webů s bohatou historií budování odkazů je potřeba využít více komplexní vyhledávání a může se stát, že můj nástroj nebude dosahovat takových kvalit jako například Web CEO.

### Podlahy-chýlek.cz

nástroj	Web CEO	GWT	Seo SpyGlass	moje řešení
celkový počet odkazů	18	1238	0	28
počet odkazujících stránek	17	-	0	20
počet odkazujících domén	10	13	0	10
počet odkazujících IP	8	-	0	7
Google Link Popularity	0	-	0	0
počet anchor textů	8	-	0	5
počet různých TLD	2	-	0	4
počet cílových URL	2	1	0	3

Tabulka 6: Srovnání výsledků pro web podlahy-chylek.cz

Z výsledků vyplývá, že u webů s nižším počtem zpětných odkazů je konečný počet odkazujících domén i stránek podobný jako u konkurenčních nástrojů (Tabulka 6). V některých případech dosahoval můj nástroj i lepších výsledků než ostatní. Nástroj Seo Spy Glass dokonce pro web podlahy-chylek.cz nenašel žádné odkazy. Také celkový čas nalezení kompletního odkazového profilu je u menšího webu znatelně kratší.

U obou projektů dopadl nejhůř Google Webmaster Tools (GWT), jehož výsledky působí dost zmateně

a neposkytují žádnou ucelenou analýzu. Nástroj Seo SpyGlass zahrnuje do celkového počtu odkazů i ty, které již neexistují, což může být na první pohled matoucí a u webů s malým počtem odkazů nepodává dokonce žádné výsledky. Nástroj Web CEO podává relevantní výsledky jak u počtu odkazujících stránek i domén, tak i u jednotlivých analýz.

Moje řešení je oproti konkurenčním lepší v aktuálnosti dat. Nově vytvořené odkazy dokáže nalézt téměř ihned po jejich indexování ve vyhledávačích. Aktuálnost profilu je podporována i tím, že systém ví, na které domény uživatel vložil odkaz a provádí tak častěji jejich kontroly. To vede k rychlejšímu nalezení nových odkazů. U jiných nástrojů musí dojít nejprve k reindexaci všech webů v databázi, aby se nové odkazy objevily. To samé platí i u odkazů, které již neexistují, nebo odkazující stránka přestala fungovat, což můj nástroj dokáže odhalit dříve.

## 6.2.2 Výsledky analýz

Bylo analyzováno několik webových projektů včetně jejich konkurentů. Všechny odkazové profily posloužily k naplnění databáze hodnocení katalogů a dalších odkazujících stránek. Agregovaná data slouží jak současným, tak i budoucím uživatelům ke zkvalitnění procesu budování zpětných odkazů. V tabulce je uveden dosavadní souhrn po provedení analýz všech vložených webů (Tabulka 7).

	Počet
Projektů	11
Analyzovaných webů	23
Nalezených odkazů	5583
Odkazujících stránek	2387
Evidovaných katalogů	2828
Ohodnocených katalogů	141

*Tabulka 7: Souhrn analýz*

Velkou roli při hledání odkazujících stránek hrála předchozí snaha o budování odkazů. Pro web, který je mladý nebo nebyl registrován do velkého množství katalogů, je snadnější a rychlejší analyzovat jeho kompletní odkazový profil. Naopak u webů s bohatou historií a velkým počtem zpětných odkazů trval celý proces mnohem déle a v porovnání s velkými konkurenty nedocházelo k tak dobrým výsledkům.

### 6.2.3 Výkonnost systému

Provedl jsem výkonnostní testy jednotlivých stahovacích úloh, abych zjistil, jak dlouho trvá analýza odkazového profilu jednoho webu. Výsledky uvádím níže v tabulce (Tabulka 8). Časy mohou být proměnlivé, protože jsou závislé na počtu nalezených odkazujících stránek a jejich obsahu.

Proces	Minut
SEOMoz API	23
Google URL	19
Google Link Popularity	15
Katalogy	29
Důležité domény	12
<b>Celkem (212 odkaz. stránek)</b>	<b>98</b>

Tabulka 8: Časy stahování

Provoz aplikace probíhá na školním serveru s parametry (Tabulka 9):

<b>Operační systém</b>	Ubuntu Linux (64-bit)
<b>Procesor</b>	Intel® Xeon® CPU E5-2660 v2 @ 2.20GHz (x4)
<b>RAM</b>	4096 MB
<b>Rychlost internetu</b>	36,88 Mbps

Tabulka 9: Parametry serveru

## 6.3 Budování zpětných odkazů

Za účelem otestování aplikace a zjištění přínosu při získávání nových odkazů jsem provedl řadu registrací do vybraných katalogů u webu **podlahy-chylek.cz**. Vybíral jsem podle jejich hodnoty, která vznikla agregací dat ze všech odkazových profilů. Naplnění společné databáze znalostí tak umožnilo řadit katalogy podle jejich skutečné kvality.

Nejprve jsem vyplnil registrační data firmy a poté přidal několik různých textových obsahů pro účely registrací. Snažil jsem se uplatnit všechny nejlepší postupy při optimalizaci, které jsem popsal v minulých kapitolách. Titulky a popisy jsem se snažil správně optimalizovat na vybraná klíčová slova.

Odkazy jsem přidával v nepravidelných počtech a v malém množství, aby to nevzbudilo podezření u vyhledávačů. Cílem bylo získat 20 zpětných odkazů během jednoho měsíce. Po úspěšné registraci se nový odkaz začal objevovat ve výsledcích vyhledávání druhý až třetí den. Jednalo se však o kvalitní katalogy a PR weby, které jsou často kontrolovány vyhledávacími roboty. Některé přidávané odkazy byly objeveny až po delší době. Jakmile se odkaz objevil ve vyhledávacích, aplikace jej byla schopna najít a zařadit do odkazového profilu.

Během procesu budování odkazů se vytvářely pravidelné reporty počtu odkazů a hodnocení webu. Pozornost byla zaměřena zejména na Page Rank a S-rank, které jsou závislé na počtu a kvalitě zpětných odkazů. Sledovány byly i pozice ve vyhledávačích na vybraná klíčová slova.

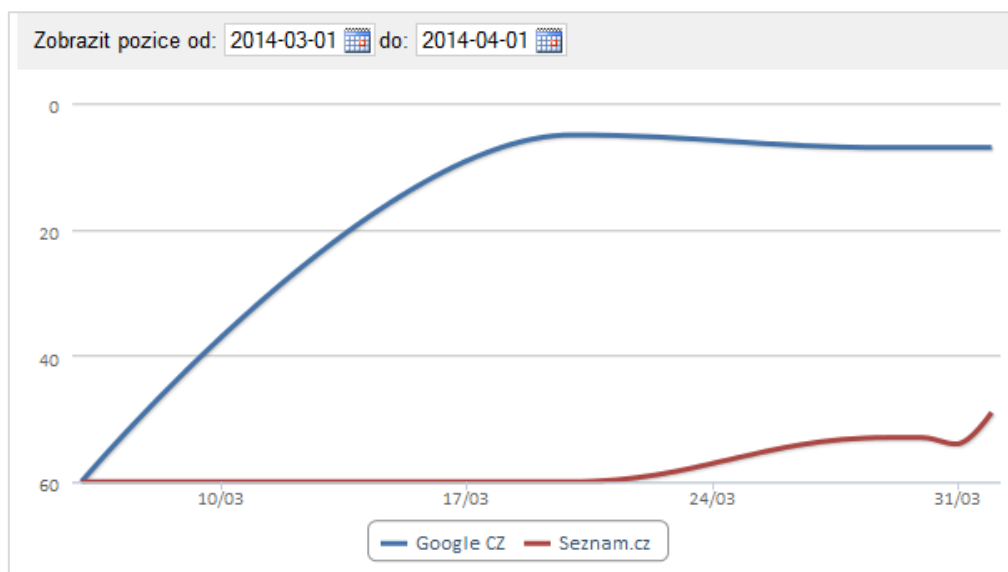
### 6.3.1 Změna hodnocení webu

Sledování a vyhodnocení změn u Page Rank a S-rank bylo obtížnější, neboť se neprojevují v krátkém časovém úseku. V průběhu budování odkazů skutečně k žádným změnám nedošlo. U webu podlahy-chylek.cz se změny projevily již po 14 dnech od poslední registrace. Konkrétně došlo ke zvýšení hodnoty S-rank z 1 na 2. Ke zlepšení aktuálního stavu stačilo pouhých 20 kvalitních odkazů během jednoho měsíce. Pro hlubší zkoumání změn v závislosti na zpětných odkazech je však potřeba dlouhý časový úsek v řádech měsíců a velké množství úsilí.

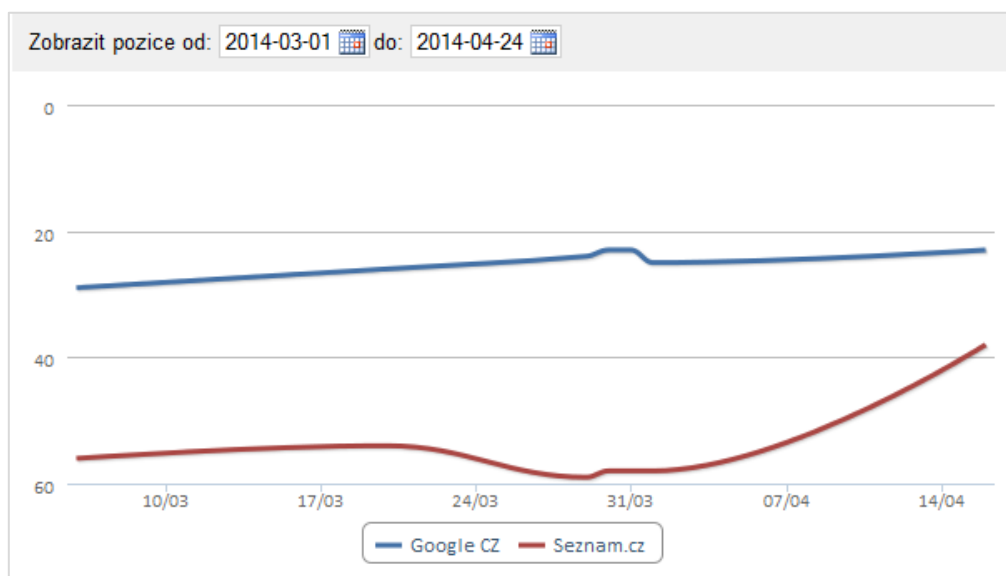
### 6.3.2 Změna pozic

V průběhu budování odkazů byly sledovány také pozice ve vyhledávačích na vybraná klíčová slova. Měření začalo ještě před samotným procesem, aby byl zaznamenán počáteční stav. Cílem bylo sledováno 16 klíčových slov souvisejících se zkoumaným webem. K měření byl použit nástroj Collabim.

Během měsíce, kdy probíhaly registrace, se pozice webu kolísavě měnily. Nebylo možné přesně určit, zda došlo ke zlepšení právě kvůli zpětným odkazům. U některých slov došlo k mírnému zlepšení, u jiných nebyla změna znatelná. Objevily se však i klíčové fráze, u kterých byla změna velmi razantní a trvalá. Například u fráze „renovace podlah“ došlo ke skokovému zvýšení z hodnot 60+ na 5 ve vyhledávači Google (Obrázek 34). U jiných se změny projevily až po skončení budování, konkrétně u fráze „pokládka podlah frýdek“ (Obrázek 35). V obou případech se jednalo o klíčové fráze, které se vyskytovaly v krátkém popisu webu používaného při budování odkazů.



Obrázek 34: Graf pozic klíčové fráze "renovace podlah"



Obrázek 35: Graf pozic fráze "pokládka podlah frýdek"

Aby byly výsledky objektivní, neprováděly se v průběhu testu na zkoumaném webu žádné úpravy obsahu ani jiné úkony, které by mohly způsobit změnu pozic. Z toho důvodu jsem testování prováděl pouze na webu, u kterého jsem tuto skutečnost mohl potvrdit.

## Konkurenční boj

Za zmínku jistě stojí i konkurenční boj mezi webovými stránkami. I to nejlepší SEO nedokáže zaručit dobré pozice, neboť konkurence provádí stejné úkony, aby se dostala na co nejvyšší pozici ve výsledcích vyhledávání. Jeden den může být web na prvním místě a druhý den třeba na čtvrtém. Řekněme, že 10 webů na první straně v SERP dělá bezchybné SEO, ale na prvním místě může být vždy jen jeden web.

## 6.4 Výsledky stahování aukcí

Výsledky stahování aukcí z portálu Aukro splnily očekávání. Pro všechny produkty v systému a jejich vygenerované vyhledávací fráze probíhaly po dobu několika měsíců pravidelné stahovací procesy. Nalezené aukce byly párovány s příslušnými produkty a pomohly k doplňkové analýze prodeje zboží. Množství vyhledávaných a nalezených položek je uvedeno v následující tabulce (Tabulka 10).

Stahování bylo spouštěno pomocí Cron úlohy a zpracování 1000 frází trvalo systému průměrně 12 minut.

Položka	Počet
<b>Produkty</b>	469
<b>Fráze</b>	10109
<b>Nalezené aukce</b>	1183
<b>Prodejci</b>	202

*Tabulka 10: Souhrn stahování aukcí*



## 7 Nápady na vylepšení

Vytvořený nástroj je sice plně funkční, ale do budoucna je možné jeho rozšíření o další funkcionality. Zlepšit by se daly i procesy vyhledávající zpětné odkazy, aby bylo dosaženo kvalitnějších výsledků.

### **Rozšíření o další zdroje**

Pro nalezení většího počtu zpětných odkazů by bylo zapotřebí přidat další zdroje. Mohou to být například vyhledávače Seznam, Bing, Centrum, Jyxo případně Yandex.

### **Vyhledávací robot**

Spuštění vyhledávacího robota (crawler), který by procházel webové stránky a ukládal si jejich odkazy do databáze, by přineslo zásadní změnu při hledání zpětných odkazů. Toto řešení používají některé velké konkurenční nástroje, ale je velmi náročné z hlediska výkonu a kapacity serveru.

### **Sociální sítě**

Zajímavé by bylo rozšíření aplikace o analýzu sociálních sítí nejen z hlediska vyhledávání odkazů, ale i jejich budování. Vytvoření funkce pro aktivní tvorbu obsahu na sociálních sítích by mohlo přinést zajímavé výsledky.

### **Výkonnější server**

Pro zrychlení analýzy odkazového profilu by bylo vhodné do budoucna umístit stahovací procesy na více serverů, kde by mohly běžet paralelně, a tak snížit celkový čas potřebný pro komplexní analýzu webu.

## 8 Závěr

V první části dokumentu jsem se věnoval popisu správných a špatných SEO technik a jejich dopadu na web. Poté jsem se podrobně věnoval budování zpětných odkazů, zejména popisu správných postupů a definování oblastí pro jejich umístění. Dále jsem zkoumal vyhledávací enginy velkých společností se zaměřením na jejich algoritmy, které hodnotí kvalitu webových stránek a udávají jejich umístění ve výsledcích vyhledávání. Zajímaly mě hlavně faktory přímo související s vyhodnocením kvality zpětných odkazů. Poznatky z této kapitoly jsem poté promítnul do vývoje vlastního nástroje na jejich měření a ověřování.

Cílem diplomového projektu bylo zjistit, které metody a postupy při budování zpětných odkazů jsou nejvhodnější a přinášejí webu nějaký užitek. Nejprve jsem nastudoval odborné materiály a články s podobnou tematikou, abych zjistil minulé i současné trendy. Zároveň jsem zkoumal konkurenční analytické nástroje se zaměřením na jejich funkce, výhody a nevýhody.

Nad rámec zadání práce byla implementace modulu pro stahování aukcí z portálu Aukro. Data z něj získaná posloužila k analýze nabídky a prodeje u mnoha produktů.

Hlavním přínosem mé diplomové práce je vytvoření univerzálního nástroje, s jehož pomocí lze analyzovat odkazový profil webu. Data z analýz provedených napříč všemi vloženými weby se agregují do společné databáze znalostí, která slouží pro doporučování stránek vhodných k budování zpětných odkazů. Aplikace zajišťuje i evidenci obsahu na odkazujících stránkách a jeho rovnoměrné použití, což značně snižuje jeho duplicitu na internetu. Při vývoji jsem vycházel z existujících řešení, podle kterých jsem implementoval základní funkce a doplnil je o řadu vlastních. Nástroj byl posléze otestován na několika reálných projektech a výsledky byly porovnány s konkurenčními řešeními.

Podařilo se odhalit skutečnou kvalitu katalogů a PR webů používaných v ČR a doporučit uživatelům ty, které mohou opravdu předat prostřednictvím odkazu nějakou hodnotu. Získaná data byla použita pro testování účinnosti odkazů z těchto katalogů. Z testu bylo zřejmé, že tyto odkazy pomohly zlepšit hodnocení zkoumaného webu a měly efekt i na pozice některých klíčových slov. Proces budování kvalitního odkazového profilu je dlouhodobou záležitostí, která stojí mnoho času a úsilí. Plný dopad všech úkonů se projeví zpravidla až po několika měsících, a proto nedokážu říct, jaké následky nastanou po mém tříměsíčním testování.

Studium této problematiky a její následné aplikování na reálných internetových projektech bylo velmi zajímavé a získal jsem při něm mnoho důležitých zkušeností, které bych chtěl využít v budoucí profesi. Nicméně se jedná o obor, který se stále vyvíjí a je zapotřebí neustále sledovat jeho vývoj. Technika, která byla efektivní před rokem, již nemusí dnes přinášet žádný užitek. Tento fakt potvrzuje i skutečnost, že v samotném průběhu psaní této diplomové práce docházelo ke změnám v trendech, ke zpřísnění vyhledávacích algoritmů a objevilo se i mnoho inovací. Myslím si, že SEO se bude i v budoucnu nadále rozvíjet a přinášet nové možnosti a trendy v optimalizaci webových stránek.

# Literatura

- [1] *Algoritmus | Seznam Náповěda*. [online]. 2014 [cit. 2014-03-16]. Dostupné z: <http://napoveda.seznam.cz/cz/hledani-fulltext-algoritmus-vyhledavani-razeni-vysledku-faq-dotazy.html>
- [2] Aktualizace Hummingbird mění Google. *Interval.cz* [online]. 2013 [cit. 2014-03-01]. Dostupné z: <http://interval.cz/clanky/aktualizace-hummingbird-meni-google/>
- [3] What to expect in SEO in the coming months. *Matt Cutts: Gadgets, Google, and SEO* [online]. 2013 [cit. 2014-03-01]. Dostupné z: <http://www.mattcutts.com/blog/what-to-expect-in-seo-in-the-coming-months/>
- [4] Penguin znovu zbrojí, je tu verze 2.0. *Firemní weblog poradenské společnosti H1.cz* [online]. 2013 [cit. 2014-03-01]. Dostupné z: <http://blog.h1.cz/aktualne/penguin-znovu-zbroji-je-tu-verze-2-0/>
- [5] What is Google Page Rank (PR)?. *Canonical SEO* [online]. 2009 [cit. 2014-03-01]. Dostupné z: <http://www.canonicalseo.com/what-is-google-page-rank/>
- [6] BRIN, Sergey a Lawrence PAGE. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *The Stanford University InfoLab* [online]. 2009 [cit. 2014-03-01]. Dostupné z: <http://infolab.stanford.edu/~backrub/google.html#pr>
- [7] DOVER, Danny a Erik DAFFORN. *SEO: optimalizace pro vyhledávače profesionálně*. Vyd. 1. Brno: Zoner Press, 2012. ISBN 978-80-7413-172-1.
- [8] KUBÍČEK, Michal. *Velký průvodce SEO: jak dosáhnout nejlepších pozic ve vyhledávačích*. Vyd. 1. Brno: Computer Press, 2008. ISBN 978-80-251-2195-5.
- [9] Link Metrics. *Mozscape Wiki* [online]. 2013 [cit. 2014-03-19]. Dostupné z: <http://apiwiki.moz.com/link-metrics>
- [10] HITS Algorithm - Hubs and Authorities on the Internet. *Www.math.cornell.edu | Department of Mathematics* [online]. 2009 [cit. 2014-03-24]. Dostupné z: <http://www.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture4/lecture4.html>
- [11] KUBÍČEK, Michal. *50 způsobů, jak získat zpětný odkaz* [online]. PRONETmedia, 2013 [cit. 2014-03-24]. ISBN 978-80-87721-01-8. Dostupné z: <http://www.zpetnyodkaz.cz/>
- [12] TOPlist - Historie. *TOPlist* [online]. 2014 [cit. 2014-04-16]. Dostupné z: <http://www.toplist.cz/stat/?a=history&type=4>

- [13] Jak funguje WebAPI. *WebAPI* [online]. 2014 [cit. 2014-04-19]. Dostupné z:  
<http://api.aukro.cz/jak-funguje-webapi/>
- [14] DoSearch. *WebAPI* [online]. 2014 [cit. 2014-04-19]. Dostupné z:  
<http://api.aukro.cz/dokumentace/vyhledavani-a-prohlizeni/dosearch.aspx>
- [15] Licenční politika. *Nette Framework* [online]. 2014 [cit. 2014-04-19]. Dostupné z:  
<http://nette.org/cs/license>

## Přílohy

### A. Adresářová struktura přiloženého DVD

Adresář	Obsah
/Implementace	Zdrojové kódy aplikace a SQL skript generující databázi
/Texty	Text práce, zadání, klíčová slova a abstrakt

## B. Datový slovník

BAC_BACKLINKSITE			
Atribut	Datový typ	NULL	Klíč
bac_backlinkSite_id	int(10) <i>Auto Increment</i>	Ne	PK
bac_backlinkSite_url	varchar(200)	Ne	Ne
bac_backlinkSite_title	varchar(150)	Ano	Ne
bac_backlinkSite_desc	varchar(350)	Ano	Ne
bac_backlinkSite_keywords	varchar(350)	Ano	Ne
bac_backlinkSite_internalLinks	decimal(10,0)	Ne	Ne
bac_backlinkSite_externalLinks	decimal(10,0)	Ne	Ne
bac_backlinkSite_pagerank	tinyint(2)	Ne	Ne
bac_backlinkSite_homepagerank	tinyint(2)	Ne	Ne
bac_backlinkSite_srank	tinyint(3)	Ne	Ne
bac_backlinkSite_alexRank	int(11)	Ne	Ne
bac_backlinkSite_linkJuice	double	Ne	Ne
bac_backlinkSite_IP	varchar(50)	Ne	Ne
bac_backlinkSite_hash	varchar(100)	Ne	Ne
BAC_BACKLINK			
Atribut	Datový typ a velikost	NULL	Klíč
bac_backlink_id	int(10) <i>Auto Increment</i>	Ne	PK
bac_backlink_linkingPage	varchar(200)	Ne	Ne
bac_backlink_linkingDomain	varchar(200)	Ne	Ne
bac_backlink_targetPage	varchar(200)	Ano	Ne
bac_backlink_anchor	varchar(200)	Ano	Ne
bac_backlink_rel	varchar(10)	Ano	Ne
bac_backlink_type	varchar(20)	Ano	Ne
bac_backlink_source	varchar(20)	Ano	Ne
bac_backlink_isCompetitor	tinyint(1) [0]	Ne	Ne
bac_backlink_datetime	datetime	Ne	Ne
cor_site_id	int(10) unsigned	Ne	FK
bac_backlinkSite_id	int(10) unsigned	Ne	FK
bac_backlink_status	varchar(10)	Ne	Ne
BAC_CATALOG			
Atribut	Datový typ a velikost	NULL	Klíč
bac_catalog_id	int(10) <i>Auto Increment</i>	Ne	PK
bac_catalog_url	varchar(50)	Ne	Ne
bac_catalog_pay	varchar(50)	Ne	Ne
bac_catalog_srank	tinyint(4)	Ne	Ne
bac_catalog_pagerank	tinyint(4)	Ne	Ne
bac_catalog_category	varchar(20)	Ne	Ne

bac_catalog_lang	varchar(20)	Ne	Ne
bac_catalog_avgLinkValue	varchar(20) [NULL]	Ano	Ne
bac_catalog_datetime	datetime	Ne	Ne
bac_catalog_xmldata	longtext	Ano	Ne
bac_catalog_lastUpdate	datetime	Ano	Ne
<b>BAC_CATALOGPLAN</b>			
<b>Atribut</b>	<b>Datový typ a velikost</b>	<b>NULL</b>	<b>Klíč</b>
bac_catalogPlan_id	int(10) <i>Auto Increment</i>	Ne	PK
bac_catalog_id	int(10) unsigned	Ne	FK
bac_lbplan_id	int(10) unsigned	Ano	FK
bac_catalogPlan_createDate	datetime	Ne	Ne
bac_catalogPlan_planDate	datetime	Ano	Ne
bac_catalogPlan_getDate	datetime	Ano	Ne
bac_catalogPlan_sendDate	datetime	Ano	Ne
bac_catalogPlan_regDate	datetime	Ano	Ne
bac_catalogPlan_publishDate	datetime	Ano	Ne
bac_catalogPlan_deleteDate	datetime	Ano	Ne
bac_catalogPlan_state	tinyint(1) [0]	Ne	Ne
cor_site_id	int(10) unsigned	Ne	FK
bac_lbPlanData_id	int(10) unsigned	Ano	FK
<b>BAC_CHECKEDSITE</b>			
<b>Atribut</b>	<b>Datový typ a velikost</b>	<b>NULL</b>	<b>Klíč</b>
bac_checkedSite_id	int(10) <i>Auto Increment</i>	Ne	PK
bac_checkedSite_url	varchar(255)	Ne	Ne
bac_checkedSite_source	varchar(10)	Ne	Ne
cor_site_id	int(10) unsigned	Ne	FK
<b>BAC_CSECREd</b>			
<b>Atribut</b>	<b>Datový typ a velikost</b>	<b>NULL</b>	<b>Klíč</b>
bac_cseCred_id	int(10) <i>Auto Increment</i>	Ne	PK
bac_cseCred_apiKey	varchar(80)	Ne	Ne
bac_cseCred_cx	varchar(80)	Ne	Ne
bac_cseCred_dayRequest	tinyint(4)	Ne	Ne
bac_cseCred_source	varchar(20)	Ano	Ne
bac_cseCred_useDate	datetime	Ano	Ne
bac_cseCred_note	varchar(50)	Ano	Ne
<b>BAC_DOWNPROGRESS</b>			
<b>Atribut</b>	<b>Datový typ a velikost</b>	<b>NULL</b>	<b>Klíč</b>
bac_downProgress_id	int(10) <i>Auto Increment</i>	Ne	PK
bac_downProgress_source	varchar(20)	Ne	Ne
cor_site_id	int(10) unsigned	Ne	FK
bac_downProgress_date	datetime	Ne	Ne

BAC_LBPLAN			
Atribut	Datový typ a velikost	NULL	Klíč
bac_lbPlan_id	int(10) <i>Auto Increment</i>	Ne	PK
cor_project_id	int(10) unsigned	Ne	FK
bac_lbPlan_name	varchar(30)	Ne	Ne
bac_lbPlan_type	varchar(10)	Ne	Ne
bac_lbPlan_createDate	datetime	Ne	Ne
cor_site_id	int(10) unsigned	Ne	FK
BAC_LBPLANDATA			
Atribut	Datový typ a velikost	NULL	Klíč
bac_lbPlanData_id	int(10) <i>Auto Increment</i>	Ne	PK
bac_lbPlanData_title	varchar(255)	Ne	Ne
bac_lbPlanData_desc	varchar(500)	Ne	Ne
bac_lbPlanData_article	mediumtext	Ano	Ne
bac_lbPlanData_type	varchar(20)	Ne	Ne
bac_lbPlanData_url	varchar(255)	Ne	Ne
cor_site_id	int(10) unsigned	Ne	FK
bac_lbPlanData_useCount	int(10) [0]	Ne	Ne
BAC_REGCOMPANYINFO			
Atribut	Datový typ a velikost	NULL	Klíč
bac_regCompanyInfo_id	int(10) <i>Auto Increment</i>	Ne	PK
bac_regCompanyInfo_name	varchar(80)	Ne	Ne
bac_regCompanyInfo_contactFirstname	varchar(30)	Ne	Ne
bac_regCompanyInfo_contactLastname	varchar(50)	Ne	Ne
bac_regCompanyInfo_ic	int(20)	Ne	Ne
bac_regCompanyInfo_dic	varchar(20)	Ano	Ne
bac_regCompanyInfo_street	varchar(100)	Ne	Ne
bac_regCompanyInfo_city	varchar(50)	Ne	Ne
bac_regCompanyInfo_zip	varchar(10)	Ne	Ne
bac_regCompanyInfo_region	varchar(50)	Ne	Ne
bac_regCompanyInfo_telephone	varchar(20)	Ne	Ne
bac_regCompanyInfo_email	varchar(50)	Ano	Ne
bac_regCompanyInfo_fax	varchar(20)	Ano	Ne
cor_site_id	int(10) unsigned	Ne	FK
cor_project_id	int(10) unsigned	Ne	FK
BAC_REPORT			
Atribut	Datový typ a velikost	NULL	Klíč
bac_report_id	int(10) <i>Auto Increment</i>	Ne	PK
bac_report_totalBacklinks	int(11)	Ne	Ne
bac_report_linkingPages	int(11)	Ne	Ne
bac_report_linkingDomains	int(11)	Ne	Ne



bac_report_linkingIP	int(11)	Ne	Ne
bac_report_googleLinks	int(11) [0]	Ne	Ne
bac_report_pagerank	tinyint(2)	Ne	Ne
bac_report_srank	tinyint(3)	Ne	Ne
bac_report_alexRank	int(11)	Ne	Ne
bac_report_fbshares	int(10)	Ne	Ne
bac_report_fblikes	int(10)	Ne	Ne
bac_report_fbcomments	int(10)	Ne	Ne
bac_report_tweets	int(10)	Ne	Ne
bac_report_linkedIn	int(10)	Ne	Ne
bac_report_gplusone	int(10)	Ne	Ne
bac_report_websiteRank	double	Ne	Ne
bac_report_date	datetime	Ne	Ne
cor_site_id	int(10) unsigned	Ne	FK
<b>BAC_WEBSITE</b>			
Atribut	Datový typ a velikost	NULL	Klíč
bac_website_id	int(10) <i>Auto Increment</i>	Ne	PK
bac_website_url	varchar(255)	Ne	Ne
bac_website_datetime	datetime	Ne	Ne
bac_website_title	varchar(255)	Ano	Ne
bac_website_desc	varchar(255)	Ano	Ne
bac_website_keywords	varchar(255)	Ano	Ne
bac_website_pagerank	tinyint(2)	Ne	Ne
bac_website_srank	tinyint(2)	Ne	Ne
bac_website_links	int(11)	Ne	Ne
bac_website_linkJuice	double	Ne	Ne
cor_site_id	int(10) unsigned	Ne	FK
cor_project_id	int(10) unsigned	Ne	FK
bac_website_type	varchar(10)	Ne	Ne
bac_website_status	bit(1) [b'1']	Ne	Ne
<b>BAC_WEBSITEPLAN</b>			
Atribut	Datový typ a velikost	NULL	Klíč
bac_websitePlan_id	int(10) <i>Auto Increment</i>	Ne	PK
bac_website_id	int(10) unsigned	Ne	FK
bac_websitePlan_createDate	datetime	Ne	Ne
bac_websitePlan_publishDate	datetime	Ano	Ne
bac_websitePlan_deleteDate	datetime	Ano	Ne
bac_websitePlan_state	tinyint(1) [0]	Ne	Ne
cor_site_id	int(10) unsigned	Ne	FK